WHOLE-GENOME BISULFITE SEQUENCING REVEALS GENOMIC REGIONS VULNERABLE TO EPIGENETIC PERTURBATION

by
Suzanne N. Martos

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
August 2017

Abstract

Environmental exposures that alter epigenetic marks can have phenotypic consequences for multiple generations. Prenatal nutrition and environmental stressors have been shown to frequently alter DNA methylation at retrotransposon-associated metastable epialleles and imprinted germline differentially methylated regions (gDMRs). Such frequent alterations suggest that these loci may be especially susceptible to epigenetic perturbation. Previously, we used whole-genome bisulfite sequencing (WGBS) of wildtype (WT) mouse embryonic stem cells (mESCs) and three mutant mESCs with altered DNA methyltransferase (DNMT) enzymes: loss of "maintenance" DNMT1 (1KO; *Dnmt1*$^{-/-}$), loss of "*de novo*" DNMT3a/3b (DKO; *Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$), and loss of "maintenance" and "*de novo*" DNMTs (TKO; *Dnmt1*$^{-/-}$/*Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$) to demonstrate a division of labor between DNMT1 and DNMT3a/3b to suppress distinct types of retrotransposons (i.e., DNMT1-maintained or DNMT3a/3b-maintained regions). Here we show that DNA methylation maintenance at imprinted gDMRs requires DNMT1 and DNMT3a/3b. Characterization of rescued-1KO mESCs (r1KO; *Dnmt1*$^{-/-}$ with exogenous *Dnmt1* cDNA) revealed that loss of methylation at known imprinted loci in 1KO cells was not rescued by exogenous DNMT1. We developed a computational method, Non-Rescued DMRs (NORED), to identify 2500 NORED regions (FDR = 0.02) genome-wide. We implemented a second computational approach, MethylMosaic, to identify 2500 genomic regions (FDR = 0.2) with bimodal read-level methylation, a characteristic that arises at gDMRs from one parental allele being fully methylated and the other being unmethylated. Overlap of the two methods, identified 207 regions, named as bimodal NORED, including 19 of 21 known gDMRs. To determine whether DNMT1-maintained or bimodal NORED regions are more susceptible to environmental perturbation than other genomic features, we compared fetal liver methylomes from male C57BL/6J mice that exposed *in utero* to Bisphenol A (BPA) through maternal supplemented diet (0 μg/kg bw/d, 10 μg/kg bw/d, or 10 mg/kg bw/d). We then identified DNA methylation-associated pathway perturbation to discover potential molecular mediators of metabolic abnormalities observed in male mice. In summary, distinct DNA methylation maintenance paradigms and epigenetic characteristics found at germline imprints and retrotransposons influenced susceptibility to altered methylation from prenatal exposure to BPA and DNA methylation-mediated pathway perturbation provided insight into adult phenotypes.

Advisor: Zhibin Wang, Ph.D.

Committee Chair: Hongkai Ji, Ph.D.

Acknowledgments

I would like to thank my advisor, Dr. Zhibin Wang for the opportunity to work on interesting projects. Dr. Wang is an innovative experimental biologist, which has provided me with a unique opportunity to develop computational approaches that take into account the underlying molecular basis of genome-wide sequencing data. I would also like to thank my current and former labmates for working directly with me on projects or providing insightful scientific discussion: Dr. Ramon Ramos, Dr. Dan Lou, Dana Freeman, Dr. Yang Gao, Dr. Teng Li, and Dr. Hongzheng Dai.

As my committee chair and collaborator, Dr. Hongkai Ji has been instrumental in providing statistical guidance so that my bioinformatic algorithms not only had a biological basis, but could also be systematically evaluated. I would like to thank dissertation readers, Dr. Winnie Tang and Dr. Anthony Leung, as well as alternate dissertation readers Dr. Kathleen Gabrielson and Dr. Mark Kohr.

I am very appreciative that I was able to collaborate with Dr. Marisa Bartolomei and Dr. Martha Susiarjo. They provided samples that enabled me to quickly translate my findings from cell culture and bioinformatic algorithms into a well-established early life exposure model. In addition to our collaborative project, Dr. Bartolomei and Dr. Susiarjo have been incredibly helpful in providing input for presentations.

I am extremely grateful to Dr. Thomas Hartung, who served on my dissertation committee, and everyone at the Center for Alternatives to Animal Testing (CAAT). From my first lab rotation working with Dr. Lena Smirnova through the past five years, everyone at CAAT has been incredibly welcoming and encouraging. In addition to Dr. Hartung and Dr. Smirnova, I would like to thank Dr. Alex Maertens, Dr. Helena Hogberg, Dr. Andre Kleensang, Georgina Harris, and Tom Luechtfeld for consistently showing their support by attending my presentations. I am particularly thankful to Dr. Maertens for her advice and thoughtful conversations regarding bioinformatic approaches in toxicology.

I am incredibly lucky to have been able to share the experience of earning a doctorate with amazing cohortmates, Dr. Makena Clive and Dr. Meena Aladdin. Getting a Ph.D. has many ups and downs and I was fortunate to be able to spend the last five years going through this process with two intelligent, thoughtful, and supportive women.

Finally, I would like to thank my family. First and foremost, I would like to thank my parents, Tammy Davenport and Martin Martos, for always encouraging me to pursue my interests. I am

immeasurably grateful to my mom for nurturing my “ability to persevere” and tendency to be assertive rather than reprimanding me for being stubborn and bossy as a child. For as long as I can remember, my dad has always told me that I could do anything and to not limit my expectations based on external biases and stereotypes. Without the influence and encouragement of my parents, this achievement would not have been possible. To my grandparents, whose hard work and dedication to family has been an inspiration. I would also like to thank my older brother, Martin R. Martos II, for motivating me to aspire to high goals and challenging me to not become complacent. To my sister-in-law, Cathryn Martos, thank you for your support and encouragement. To my neice, Emma Suzanne, your spunky and inquisitive personality reminds me of how exciting it is to be able to constantly learn new things. And last but not least, my significant other, Dr. Richard Allen, who warned me, “you don’t want to do it,” when I was considering returning to school for my doctorate. Nonetheless, he has been infinitely accommodating, emotionally supportive, and gracious enough to limit saying, “I told you so” to at most once per week.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Part A–Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease

(*Originally published as:* Martos S.N., Tang W-Y, Wang Z. Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. Prog Biophys Mol Biol. Elsevier Ltd; 2015 Jul;118(1-2):44–54. PMID: 25792089)

Abstract

Epigenetic mechanisms involving DNA methylation, histone modification, histone variants and nucleosome positioning, and noncoding RNAs regulate cell-, tissue-, and developmental stage-specific gene expression by influencing chromatin structure and modulating interactions between proteins and DNA. Epigenetic marks are mitotically inherited in somatic cells and may be altered in response to internal and external stimuli. The idea that environment-induced epigenetic changes in mammals could be inherited through the germline, independent of genetic mechanisms, has stimulated much debate. Many experimental models have been designed to interrogate the possibility of transgenerational epigenetic inheritance and provide insight into how environmental exposures influence phenotypes over multiple generations in the absence of any apparent genetic mutation. Unexpected molecular evidence has forced us to reevaluate not only our understanding of the plasticity and heritability of epigenetic factors, but of the stability of the genome as well. Recent reviews have described the difference between transgenerational and intergenerational effects; the two major epigenetic reprogramming events in the mammalian lifecycle; these two events making transgenerational epigenetic inheritance of environment-induced perturbations rare, if at all possible, in mammals; and mechanisms of transgenerational epigenetic inheritance in non-mammalian eukaryotic organisms. This paper briefly introduces these topics and mainly focuses on (1) transgenerational phenotypes and epigenetic effects in mammals, (2) environment-induced intergenerational epigenetic effects, and (3) the inherent difficulties in establishing a role for epigenetic inheritance in human environmental disease.

1. Introduction

Barker et al. postulated that organs undergo developmental programming *in utero* that predetermines subsequent physiological and metabolic adaptations during adult life (Barker et al., 1993; Hales and Barker, 2001). Classic examples include association between low birth weight and a greater risk of coronary heart disease, hypertension, stroke, depression, type 2 diabetes, and osteoporosis in later life (Barker et al., 1993; Fernandez-Twinn and Ozanne, 2006; Gluckman and Hanson, 2004). The observation from the Dutch famine studies also provided a proof of this concept. The paradigm is rooted in the process of developmental plasticity (Bateson et al., 2004) that argues that most human organs, prior to full maturation, are capable of re-directing their course of development based on early life clues that forecast later-life demands. It is thought that a fetal environment can alter various organ systems, rendering the individual more susceptible to disease in later life when there is a chance to encounter a second, disease-promoting stimulus (Burdge et al., 2008). Epigenetics now underpins the developmental reprogramming by demonstrating the molecular relationship between the environment and gene expression (Jirtle and Skinner, 2007; Tang and Ho, 2007).

Within the last decade, studies have been published, suggesting the possibility of transgenerational epigenetic effects in mammals (Anway et al., 2005; Padmanabhan et al., 2013; Rassoulzadegan et al., 2006). These observations force us to question whether environment-altered epigenetic marks can be inherited over generations to contribute to human disease. We now present the possible epigenetic mechanisms underlying the developmental plasticity (adaptive epigenetic variations) and epigenetic inheritance towards exogenous environmental factors and whether these epigenetic effects persist in subsequent generations. We briefly define transgenerational, discuss how mammalian epigenetic reprogramming should prevent most epigenetic marks from persisting across multiple generations, and summarize established mechanisms of transgenerational epigenetic inheritance in non-mammalian eukaryotic organisms. Then we review mammalian models demonstrating transgenerational phenotypes and environment-induced intergenerational epigenetic effects. Finally, we discuss the difficulties in establishing a role for epigenetic inheritance in human environmental disease and why this necessitates an

understanding of the molecular basis underlying transgenerational phenotypes to determine potential implications for human health.

1.1. Transgenerational versus intergenerational epigenetic effects

Many studies have described intergenerational effects as evidence for transgenerational inheritance. However, correcting the misconception that these two terms are interchangeable is necessary for progress to be made in the study of epigenetic inheritance in human disease. Clarifying these definitions will allow researchers to develop and choose appropriate model systems for examining environmental influences on intergenerational and transgenerational phenotypes and associated mechanisms. In mammals, pregnant females exposed to environmental factors, such as nutrition, hormones, toxicants, or stress can affect fetal development. Such *in utero* exposures can also affect developing germ cells within the fetus (Skinner, 2008). Environment-induced epigenetic changes are referred to as intergenerational when they occur in the adult female organism (F0), the first generation of offspring (F1), or the second generation of offspring (F2), because the adult, the fetus, and the primordial germ cells (PGCs) would be directly exposed to the inducing agent. Effects may be transgenerational only when observed in subsequent generations (F3 or later) in the absence of exposure to the inducing agent or environmental factor that initiated the change. Effects observed in the male germline during the second-generation offspring (F2) may be transgenerational when induced during exposure to the adult male (F0) and his germline (F1). Importantly, this does not imply that all epigenetic effects in F3 after gestational female exposure or F2 after male exposure are necessarily epigenetic inheritance. Parental effects (Daxinger and Whitelaw, 2012; Whitelaw and Whitelaw, 2008), recapitulation (Waterland, 2014) and DNA sequence changes (Heard and Martienssen, 2014) should be excluded. For example, that seminal fluid can affect the uterine environment (Bromfield, 2014; Robertson, 2005) and impact offspring phenotype (Bromfield et al., 2014) implies that paternal effects could also influence developing PGCs (F2), independent of germline-transmitted effects. Examples of non-germline maternal effects are described later in Sections 2.2 and 2.4. Several reviews have previously described distinguishing between intergenerational and transgenerational effects in greater detail (Daxinger and Whitelaw, 2012; Heard and Martienssen, 2014; McCarrey, 2014; Schmidt, 2013; Skinner, 2013). Up to date, the majority of environmental toxicants are shown to influence somatic cells (in

F0 and/or F1 germ cell) via epigenetic mechanisms and induce disease phenotypes in mammals but not transmit those epigenetic effects into F3 (mother exposed) or F2 (father exposed). Transgenerational inheritance of epigenetic changes is commonly shown in plants only. Limited studies have demonstrated that environmental toxicants are able to promote transgenerational inheritance of phenotypes and diseases states in mammals. Findings from either aspect can help us to define the exposure window to the nutritional, hormonal, or stress/ toxin environments that may induce the adaptive and/or heritable epigenetic changes on the developing embryo and its germline, and cause disease phenotypes in subsequent generations.

1.2. Epigenetic reprogramming in mammals

An understanding of the resetting of epigenetic marks during development is needed to investigate the role of epigenetic inheritance in human disease. Within the mammalian life-cycle, the genome undergoes two global epigenetic reprogramming events, once in the zygote and second in the developing PGCs, reviewed in (Cowley and Oakey, 2012; Hackett and Surani, 2012; Heard and Martienssen, 2014; McCarrey, 2014). For zygote reprogramming after fertilization, the paternal genome is rapidly demethylated, and the maternal genome is passively demethylated; after implantation, genome-wide *de novo* methylation occurs and is completed by embryonic day 6.5 (E6.5) in mice (Smith et al., 2012). Regions of the mouse genome resistant to zygotic reprogramming include imprinted differentially methylated regions (DMRs), intracisternal A particles (IAPs), and L1Md_A retroelements, a family of long interspersed elements (LINEs) (Smith et al., 2012). Coincident with post-fertilization demethylation, the mammalian zygote undergoes a process, maternal-to-zygotic transition, during which maternal RNAs are degraded, and the embryonic genome becomes transcriptionally active. The timing of zygote genome activation varies across species (1–2-cell stage in mice and 4–8-cell stage in humans) and prior to this, pre-ovulation-accumulated maternal RNAs and proteins direct developmental processes, reviewed in (Li et al., 2013; Tadros and Lipshitz, 2009).

Germline reprogramming has indeed raised the question as to how these ‘reprogrammed’ marks are being inherited. Gamete imprinting via DNA methylation machinery is reported as establishment of epigenetic patterns for future generations (Trasler, 2006). Following post-implantation *de novo*

methylation, cells in the epiblast are induced to become PGCs and global demethylation occurs (E6.5 to E9.5); a subset of specific sequences, including imprinted DMRs, maintain methylation until E10.5 and are then demethylated by E13.5 (Seisenberger et al., 2012). After E13.5 sex-specific *de novo* methylation occurs: in males, methylation of germ cells is completed by birth; in females, methylation occurs between birth and puberty (Cowley and Oakey, 2012; Smallwood et al., 2011). At E13.5, mouse PGCs are extremely hypomethylated, but site-specific resistance to demethylation occurs for some IAPs, as well as single-copy genomic regions including, some CpG islands (CGIs) and non-CGI promoters near IAPs (Seisenberger et al., 2012). A recent study identified 4730 regions, including 233 single-copy regions, that maintained greater than 40 percent CpG methylation on E13.5 in female mouse PGCs (Hackett et al., 2013b; Waterland, 2014). That "two rounds of epigenetic erasure leave little chance for inheritance of epigenetic marks, whether programmed, accidental, or environmentally induced," argues against transgenerational epigenetic inheritance in mammals (Heard and Martienssen, 2014). Indeed, for transgenerational epigenetic inheritance to occur, epigenetic factors would have to either evade or otherwise be protected from global erasure/degradation during reprogramming. More studies on the underlying mechanisms of how environmental factors modulate the epigenetic modifications in germ cells are required in the future. The reversible and responsive nature of epigenetic modifications toward the environment leads us to wonder whether this reprogramming will persist or be further modified throughout the course of life of the organism.

1.3. Transgenerational epigenetic inheritance in eukaryotic organisms

Mechanisms of transgenerational epigenetic inheritance have been demonstrated mostly in plants and non-mammalian animal models. Transcriptional silencing in plants and non-mammalian animal models indicate that small RNA may initiate stable chromatin silencing that can then be inherited over generations (Heard and Martienssen, 2014). Often associated with transposable elements or repeats, RNA-Dependent DNA Methylation (RdDM) and paramutation, in plants, depend on RNA interference (RNAi), reviewed in Heard and Martienssen (2014). Plant epigenetic reprogramming and epiallele transmission are discussed in (Kawashima and Berger, 2014). Described as a "conversion-type phenomenon" (Coe, 1959) to explain the

heritable and reversible induction of phenotype (Brink, 1956), paramutation involves one allele inducing a heritable epigenetic change in the other without DNA sequence modification (de Vanssay et al., 2012).

Paramutation in the fruit fly, *Drosophila melanogaster*, establishes transgenerational silencing of transposable elements. Silencing is mediated by Piwi-interacting RNAs (piRNAs), small RNAs that interact with Argonaute-family proteins in the Piwi clade (Stuwe et al., 2014). Transcribed from transposon sequence-enriched genomic regions, termed piRNA clusters, primary piRNAs are then processed and incorporated into Piwi proteins to silence target sequences (Brennecke et al., 2007). In addition to recognizing established transposable elements, the piRNA pathway can recognize new transposable elements that invade the host genome. Insertion of an exogenous sequence into a piRNA cluster results in repression of similar sequences in euchromatic regions of the genome (Muerdter et al., 2012; Roche and Rio, 1998; Ronsseray et al., 2003) and novel transposons can become integrated into piRNA clusters once an organism has been exposed (Khurana et al., 2011). Stable transgenerational epigenetic inheritance (over 50 generations) via paramutation is initiated when maternally inherited piRNAs, transmitted through oocyte cytoplasm, lead to *de novo* silencing of homologous regions on the paternal allele (de Vanssay et al., 2012).

In the roundworm, *Caenorhabditis elegans*, dsRNA or piRNA can initiate epigenetic silencing via the RNAi/chromatin pathway, which involves RNAi factors, chromatin proteins, and histone methyltransferases (Heard and Martienssen, 2014). Epigenetic silencing at endogenous genes and transgenes can last more than 20 generations and depends on *rrf1*, an RNA-dependent RNA polymerase (Heard and Martienssen, 2014). Additional details for small RNA-mediated transgenerational inheritance in *C. elegans* are reviewed in (Castel and Martienssen, 2013).

Many of these established pathways and factors implicated in transgenerational epigenetic inheritance in non-mammalian models have not been thoroughly explored in mammalian models. Studies attempting to identify non-genetic inheritance in mammalian models should expand the scope of their studies to include similar potential mechanisms.

2. Transgenerational phenotypes in mammals

The discovery that vinclozolin, an anti-androgenic fungicide classified as nonmutagenic, could produce transgenerational (F1 through F4) effects on male fertility and altered DNA methylation patterns in F2 and F3 sperm (Anway et al., 2005) lead to the hypothesis that environment-induced epigenetic changes, inherited through the germline, could contribute to disease. Since then, a number of environmental chemicals have been demonstrated to induce transgenerational phenotypes. In spite of this, a clearly defined molecular basis for inheritance, presumed to be epigenetic, has yet to been established (Schmidt, 2013). The proposed mechanism is that "epimutations" induced in the male germline become "imprinted-like" allowing them to escape post-fertilization DNA methylation erasure; the altered epigenome then leads to altered transcriptomes in somatic cells resulting in adult-onset diseases (Nilsson and Skinner, 2014; Skinner et al., 2013a). However, DNA sequence changes cannot be ruled out when genes responsible for phenotypes are unknown (Heard and Martienssen, 2014). Here we review the studies demonstrating transgenerational phenotypes and associated epigenetic changes induced by environments.

2.1. Transgenerational effects of environmental exposures

In the initial study of vinclozolin (Anway et al., 2005), gestating female (F0) rats were exposed to vinclozolin between days E8 and E15, corresponding with epigenetic reprogramming of germs cells and sex determination. Following transient gestational exposure to F0 females, F1, F2, F3, and F4 adult males had significantly increased spermatogenic apoptosis, decreased sperm number, and reduced sperm motility. Similar effects on sperm were observed in offspring of vinclozolin-lineage F2 males crossed with untreated females, but not in vinclozolin-lineage F2 females crossed with untreated males, suggesting that the phenotype is transmitted through the male, but not female parent. PCR amplification of methylation-sensitive restriction enzyme digested DNA demonstrated altered DNA methylation patterns in F2 (4 out of 8) and F3 (2 out of 5) vinclozolin-treated lineages. The transgenerational increase in spermatogenic cell apoptosis was replicated in inbred Fisher and outbred Sprague Dawley rat strains (Anway et al., 2006a). Subsequent studies found additional adult disease phenotypes after female F0 vinclozolin exposure, including prostate abnormalities in F1–F4 males (Anway and Skinner, 2008), severe anemia during pregnancy in F1–F3 females (Nilsson et al., 2008), and sexually dimorphic changes in anxiety behaviors

(Skinner et al., 2008). These studies demonstrate that transgenerational manifestation of vinclozolin-induced phenotypes depends on the sex of the offspring and originate from exposure during specific developmental intervals. Additionally, transcriptome characterizations found mRNA expression of 196 genes to be consistently different in testis among the F1–F3 vinclozolin-lineage animals compared to the F1–F3 controls (Anway et al., 2008). Affected genes had roles histone modification, chromatin remodeling, and DNA methylation, including significant reductions of *Dnmt3a* in F1–F2, *Dnmt1* in F1–F3, *Dnmt3L* in F1–F3, and *Ehmt1* in F1–F3 (Anway et al., 2008). Altered expression of these epigenetic enzymes provides the possibility that epigenetic processes could be involved. Transcriptome alternations in prostate (F3 males) (Anway and Skinner, 2008), Sertoli cells (Guerrero-Bosagna et al., 2013), and hippocampus and amygdala (F3 males and females) (Skinner et al., 2008) have also been reported.

In contrast, a different group exposed outbred female (F0) Wistar rats to vinclozolin during gestation (E6 to E15) and reported no adverse transgenerational anti-androgenic effects in F1–F4 male offspring for two separate studies (SCHNEIDER et al., 2008; Schneider et al., 2013). The first study reported decreased and the second reported increased spermatogenic cell apoptosis in F1–F3 male offspring after maternal F0 vinclozolin treatment. A study by an additional independent research group, using inbred Sprague-Dawley rats, reported no effects on spermatogenesis in F1 or F2 males after maternal (F0) exposure to vinclozolin (Inawaka et al., 2009). The reason is undetermined, but the discrepancy suggests the possible influence of genetic variations among strains (Heard and Martienssen, 2014; Inawaka et al., 2009; Schneider et al., 2013). Additionally, the original research group (Anway et al.) reported intraspecies differences in susceptibility to vinclozolin in mouse strains after F0 female exposure during gestation (E7 to E13). Inbred 129 mice were not sensitive to vinclozolin-induced prostate, kidney, or testis phenotypes, but apoptosis of spermatogenic cells occurred; whereas, outbred F3-generation CD-1 mice developed prostate, kidney, and testis abnormalities (Guerrero-Bosagna et al., 2012). A second group exposed inbred FVB/N mice to vinclozolin during gestation (E10 to E18) and reported changes in CpG methylation at five imprinted DMR in sperm; paternally imprinted genes decreased and maternally imprinted genes increased (Stouder and Paoloni-Giacobino, 2010). Methylation changes were observed in F1 through F3 but appeared to be returning to normal methylation levels by F3. Sperm motility was reduced in F1, but not in F2 and F3. Although not a genome-wide analysis, this suggests that CpG methylation changes at imprinted genes were

not correlated with phenotype (Schneider et al., 2013). Alternatively, these results may not be comparable to the previously described studies because exposure was extended to E18, which overlaps with the androgen receptor vulnerability window (Schneider et al., 2013) and, therefore, was not limited to germ cell reprogramming and sex determination. Herein, it suggests that the experimental designs and exposure windows determine the transgenerational effect of vinclozolin in epigenome.

Since the initial vinclozolin study, a number of endocrine-disrupting chemicals (EDCs), including 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (Manikkam et al., 2012b), permethrin and N,N-diethyl-meta-toluamide (DEET) mixture (Manikkam et al., 2012c), jet fuel JP-8 (Tracey et al., 2013), plastics mixture (bisphenol-A (BPA), bis (2-ethylhexyl)phthalate (DEHP), dibutyl phthalate (DBP)) (Manikkam et al., 2013), dichlorodiphenyltrichloroethane (DDT) (Skinner et al., 2013b), and methoxychlor (Manikkam et al., 2014) have been assessed for phenotypes and DNA methylation changes in F3 rat sperm using similar designs as the original vinclozolin study. Transgenerational phenotypes observed in rats include prostate and kidney disease (Anway and Skinner, 2008; Anway et al., 2006b; Manikkam et al., 2012b; 2012c; 2012a; 2013; Tracey et al., 2013), mammary tumors (Anway et al., 2006b), abnormalities of the immune system (Anway et al., 2006b; Tracey et al., 2013), neurological and behavioral effects (Crews et al., 2012; 2007; Gillette et al., 2014; Skinner et al., 2008), reproductive effects (Guerrero-Bosagna et al., 2013; Nilsson et al., 2008; 2012), altered mate preference (Skinner et al., 2014), and obesity (Skinner et al., 2013b; Tracey et al., 2013), reviewed in (Crews et al., 2014; Nilsson and Skinner, 2014). In addition to the paternally-transmitted phenotypes, recent studies demonstrate that some effects from DDT and methoxychlor are transmitted maternally (Manikkam et al., 2014; Skinner et al., 2013b). Independent research groups have also reported transgenerational phenotypes from exposure to EDCs or other environmental factors. Effects observed in mice from gestational (F0) exposure were altered gene expression, social interaction, and social recognition (Wolstenholme et al., 2012; 2013) from BPA; adipose depot size, increased adipocytes, and fatty livers from tributyltin (TBT) (Chamorro-García et al., 2013); spermatogonial stem cell defects from di-(2-ethylhexyl) phthalate (DEHP) (Doyle et al., 2013); and hyperactivity from nicotine (Zhu et al., 2014). Taken together, these studies suggest that vinclozolin is not unique in its ability to induce transgenerational effects and that gestational exposure can lead to a diverse array of phenotypic consequences in subsequent generations.

Epimutations have been proposed as potential evidence for transgenerational epigenetic inheritance of toxicant-induced disease phenotypes (Nilsson et al., 2012). DMR, considered to be epimutated if reproducible in three experiments, in gene promoters were determined in F3 male rat sperm: 48 for vinclozolin (Guerrero-Bosagna et al., 2010), 50 for dioxin, 33 for jet fuel (Manikkam et al., 2012a), 197 for plastics mixture, 361 for pesticides mixture, 39 for DDT, and 37 for methoxychlor (Manikkam et al., 2014). The locations of the epimutations appear to be unique to the chemicals as there is little overlap between most pair-wise comparisons and no overlap among all chemicals; plastics mixture and pesticides mixture had the most overlap, with 109 epimutations in common (Manikkam et al., 2014). Sperm from F3 vinclozolin-lineage CD-1 mice had 66 promoters with 68 differentially methylated regions (Guerrero-Bosagna et al., 2012). These results indicate that the effect a toxicant has on the epigenome exhibits some chemical specificity.

The window of susceptibly for induction of epimutations in the rat, E7/8 to E14/15, is equivalent to E6/7 to E12/13 in the mouse, which coincides with global demethylation in PGCs and hypomethylation, but excludes the remethylation phase (Section 1.2) (McCarrey, 2014). Differentially expressed genes and epimutations were determined in germ cells from male F3 vinclozolin-lineage rats on E13 (PGCs) and E16 (prospermatogonia) (Skinner et al., 2013a). There were 592 differentially expressed genes between vinclozolin-lineage germ cells and controls at E13 and 148 at E16, with only 25 at both E13 and E16 (Skinner et al., 2013a). E13 PGCs had 24 epimutations and E16 prospermatogonia had 13 epimutations, with only one gene promoter in common; neither E13 nor E16 cells had epimutations in common with those previously identified in male F3 vinclozolin-lineage mature sperm. This finding does not support specific CpG methylation-based epimutation(s) being programmed in PGCs and directly transmitted from sperm to offspring for the inheritance of vinclozolin phenotypes. Unfortunately, there are disadvantages in using a rat model to study potential epigenetic mechanisms involved in the transgenerational inheritance of the observed disease phenotypes. Extrapolating from findings in mice (Section 1.2), we assume that reprogramming-resistant regions could exist in the rat genome as well. Determining whether rat epimutations would be resistant to post-fertilization reprogramming is hindered in that much less is known about the rat genome and epigenome, compared to the mouse and human genomes. For example, genome-wide, base-resolution DNA methylation profiling across different stages of embryonic and germ cell

development have not yet been published for the rat. Furthermore, genomic regions resistant to germline erasure in the mouse are often associated with transposable elements or repeats and likely display interspecies differences in position and regulation, which prevents direct extrapolation of resistant loci from the mouse genome to the rat genome (Sections 1.2, 2.4 and 4.3).

Overall, attributing these transgenerational effects on disease susceptibility to non-genetic mechanisms is supported by differential DNA methylation observed in sperm of the F3 generation following gestational exposure of F0 females to nonmutagenic toxicants compared to unexposed controls. Perturbations to epigenetic regulatory enzymes suggest that these phenotypes involve epigenetic processes. Currently, there is not enough detailed molecular evidence to demonstrate the transgenerational maintenance of toxicant-induced epigenetic status (i.e., an epigenetic mark altered by exposure that is present in the offspring and inherited from the parental germline (Waterland, 2014)) or to conclusively rule out confounders, such as DNA sequence changes, parental effects, or recapitulation. For these reasons, the specific mechanisms of heritability for these toxicant-induced transgenerational effects need to be elucidated. Research is also needed to clarify how the epimutations affect gene regulation and disease etiology (Nilsson and Skinner, 2014).

2.2. Mutation-induced epigenetic instability and transgenerational phenotypes

A recent mouse study demonstrated the importance of accounting for parental effects in experimental models for exploring mechanisms of transgenerational inheritance in mammals. In the study, they found that the disruption of folate metabolism caused by a mutation in the methionine synthase reductase (*Mtrr*) gene of F0 females or males caused epigenetic perturbations and lead to growth defects and congenital malformations (Padmanabhan et al., 2013). Effects persisted through generations F2–F4 of wild-type offspring derived from wild-type mothers (F1). F0 metabolic deficiencies could potentially alter reproductive phenotypes in F1 wild-type mothers (e.g., affect F1 uterine development). If the F1 (wild-type) maternal environment were responsible for the observed adverse developmental outcomes and epigenetic perturbations, then F3 germ cells developing in F2 fetuses would be considered exposed to the altered maternal environment (F1 phenotype) caused by the *Mtrr* mutation in the F0 generation. To separate potential maternal effects from those directly caused by F0 grandparental *Mtrr*-deficiency on F2

gametes, embryo transfer experiments were conducted on grandprogeny (F2). Results from these experiments implicated maternal uterine environment in growth defects and maternal grandparental *Mtrr*-deficiency in transgenerational congenital malformations. Affected wild-type grandprogeny (F2) derived from *Mtrr*-deficient F0 males or females were determined to have epigenetic instability, indicated by alterations in DNA methylation at imprinted DMRs in placentas; whereas, phenotypically normal wild-type litters (F1) from *Mtrr*-deficient F0 males did not display epigenetic instability. Taken together this suggests some phenotypic effects initiated by maternal grandparental *Mtrr* deficiency are transmitted through the germ cells and are, at least in part, mediated by epigenetic mechanisms.

2.3. RNA-mediated paramutation-like effects

In addition to toxicant- and mutation-induced transgenerational effects, RNA-mediated, paramutation-like effects have been described in mice. In this example, wild-type offspring derived from a parent (male or female) heterozygous for an engineered, mutated allele, Kit^{tm1Alf}, displayed the mutant phenotype, white feet and tail tips (Rassoulzadegan et al., 2006). The phenotype was still partially visible in F2 but disappeared gradually in subsequent generations. The inheritance of the phenotype, despite the absence of genetic inheritance of the Kit^{tm1Alf} allele, suggests the possibility of epigenetic inheritance. Microinjection of RNA isolated from Kit^{tm1Alf} heterozygote sperm or brain, miR-221, or miR-222 into fertilized eggs was sufficient to induce the phenotype (Rassoulzadegan et al., 2006), which indicates RNA-mediated effects. The transgenerational inheritance does not occur in other *Kit* null alleles without a transgene insertion (Daxinger and Whitelaw, 2012), which raises the question of whether similar phenomena would arise naturally. The same group demonstrated that the injection of other microRNAs into fertilized eggs results in the transgenerational inheritance of additional phenotypes but did not lead to a stable increase in microRNAs to pass to the next generation (Grandjean et al., 2009; Wagner et al., 2008). This observation suggests that the transmission to subsequent generations is initiated, but not maintained, by the same RNA molecules. Furthermore, establishment of the paramutant-like phenotypes requires *Dnmt2*, responsible for methylation of cytosine in RNA (Kiani et al., 2013). While initiated by the transfer of RNA from sperm to offspring, the molecular basis for how the observed phenotypes are inherited

throughout generations has not been established. Importantly, these studies establish that RNA-mediated epigenetic mechanisms might play a similar role in mammalian transgenerational inheritance, as they do in plants and non-mammalian animal models.

2.4. Transgenes and metastable epialleles

The non-mendelian inheritance of transcription activity at transgenes is presumed to be epigenetic because inbred littermates have variable transcription (Daxinger and Whitelaw, 2012). Variable expression and non-mendelian inheritance patterns of endogenous genes at metastable epialleles, defined as "allele[s] at which the epigenetic state can switch and establishment is a probabilistic event" (Rakyan et al., 2002), are often associated with transposable elements and are thought to result from incomplete reprogramming of epigenetic marks (Rakyan et al., 2001).

For example, the agouti locus (*A*) controls pigmentation of the mouse, the coat color phenotype of A^{vy}/a heterozygotes ranges from yellow to various degrees of yellow background with black mottling to pseudoagouti (i.e., resembles wild-type agouti pattern) (Dickies, 1962). Coat color correlates with other varied phenotypes observed in genotypically similar littermates, including obesity and susceptibility to tumor formation (Wolff, 1978). Furthermore, the maternal and grandmaternal, but not the paternal, phenotype affects the distribution of phenotypes among offspring derived from A^{vy}/a heterozygotes crossed with *a/a* homozygotes (Morgan et al., 1999; Wolff, 1978), suggesting that phenotypic variation is mediated by maternal factors. The A^{vy} mutation, caused by an IAP element insertion 100 kb upstream of the transcription start site for the wild-type coding sequence, drives ubiquitous, constitutive transcription of a gene that has tissue- and stage-specific expression in wild-type mice (Bultman et al., 1992; Duhl et al., 1994; Waterland and Jirtle, 2003). Varying degrees of epigenetic silencing, indicated by hypermethylation at the IAP of the A^{vy} locus, suppresses transcription of agouti gene to generate mottled and pseudoagouti phenotypes (Morgan et al., 1999).

Physiologic and metabolic differences among mothers were thought to explain dependence of offsprings' phenotype on maternal phenotype (Wolff, 1971; 1978). Embryo transfer experiments (to

separate maternal environment from germline effects) indicate that phenotype distribution results from inheritance, not maternal environment (Morgan et al., 1999). Interestingly, characterization of CpG methylation in gametes, zygotes, and blastocysts demonstrated that the A^{vy} allele undergoes dynamic post-fertilization DNA methylation reprogramming consistent with established genome-wide patterns (Blewitt et al., 2006). That is paternal A^{vy} allele was dramatically reduced (from 73% to 15%) within 9–11 h after fertilization, indicating active demethylation; maternal A^{vy} allele was not rapidly demethylated (69% methylation in zygote), but was completely unmethylated by the blastocyst stage (Blewitt et al., 2006). Thus, CpG methylation at the A^{vy} allele is not likely the epigenetic mark responsible for the transgenerational phenotypes (Daxinger and Whitelaw, 2012), because it is erased within the mammalian lifecycle. In A^{iapy} mice, which are similar to A^{vy} mice in that DNA methylation at an IAP insertion inversely correlates with agouti expression and determines the distribution of offspring phenotypes (Michaud et al., 1994), low expression of maternal oocyte-specific *Dnmt1* isoform shifted offspring coat distribution toward yellow (Gaudet et al., 2004). This result indicates that the pre-implantation control of agouti expression contributes to adult phenotype (Gaudet et al., 2004); however, whether the A^{iapy} epiallele escapes post-fertilization reprogramming or whether the A^{vy}/a phenotypes are similarly dependent on oocyte-specific *Dnmt1* has not been reported.

3. Environment-induced intergenerational epigenetic effects

Currently, studies showing transgenerational epigenetic inheritance towards environmental stimuli are limited. Nonetheless, a number of human and animal studies have demonstrated that toxicant exposure and nutritional status during pregnancy can affect epigenetic marks in developing offspring, which provides a molecular explanation for increased disease susceptibility from fetal exposures (Bailey and Fry, 2014; Hilakivi-Clarke, 2014; O'Hagan and Tang, 2013; Soubry et al., 2014; Waterland, 2014). The Agouti viable yellow (A^{vy}) mouse model is, perhaps, the most well characterized animal model demonstrating a role for environmental factors in phenotypic variation through alteration of DNA methylation. Methyl supplementation of female (F0) diet during gestation shifted the distribution of offspring (F1) (Wolff et al.,

1998) and grand offspring (F2) (Cropley et al., 2006) phenotypes toward pseudoagouti when the A^{vy} allele is inherited from F0 males to F1 females to F2 offspring, indicating intergenerational effects from F0 gestational exposure. Gestational methyl supplementation-induced phenotypic changes correlated with the degree of CpG methylation at the IAP insertion site (Waterland and Jirtle, 2003). Similarly, maternal methyl supplementation increased CpG methylation of the $Axin^{Fu}$ metastable epiallele and shifted the phenotype distribution of offspring (Waterland et al., 2006). Environment-mediated epigenetic regulation of the mutant allele(s) provides a molecular basis for the phenotypic variation observed among these isogenic mice.

Other environmental exposures during gestation have also been shown to affect offspring phenotype and CpG methylation of the A^{vy} epiallele. Genistein, a soy phytoestrogen, induced hypermethylation and shifted offspring distribution toward pseudoagouti (Dolinoy et al., 2006). Bisphenol A (BPA), an estrogenic chemical used to manufacture plastics and resins, induced hypomethylation and shifted offspring distribution toward yellow, but methyl donor or genistein supplemented maternal diets were able to counter the effects of BPA on coat color distribution (Dolinoy et al., 2007b). However, methyl donor-induced hypermethylation, indicated by coat color, was not inherited transgenerationally through the female germline (Waterland et al., 2007).

Animal models involving dietary manipulation have demonstrated the importance of gestational nutrition in determining offspring disease susceptibility. *In utero* exposure to high-fat diet in mice is shown to induce paternal obesity with increase in adiposity and initiate intergenerational transmission of obesity and insulin resistance in two generations of offspring accompanied with changes in sperm microRNA content and germ cell methylation status (Fullston et al., 2013). In contrast, caloric restriction during pregnancy alters DNA methylation in the sperm of F1 offspring mice at regions resistant to zygotic reprogramming. Although these changes do not persist in F2 tissues, the results suggests *in utero* nutritional exposures during critical windows of germ cell development can impact the male germline methylome that may influence the risk of metabolic disease in the offspring (Radford et al., 2014). Nonetheless, another group demonstrated no changes in male germline DNA methylation in a dietary protein restriction model

(Carone et al., 2010). Overall, it is clear that both toxicants and nutrients can induce epigenetic changes with intergenerational phenotypic consequences. Certainly, in animal models, the exposure window to the environmental stimuli is critical for the establishment of intergenerational epigenetic changes.

4. Establishing a role for epigenetic inheritance in human environmental disease

4.1. Transgenerational human disease

Evidence of transgenerational epigenetic inheritance in plant and animal models suggests that it may be a ubiquitous phenomenon in eukaryotic organisms and could plausibly occur in mammals. Evidence from mammalian animal studies indicates that epigenetic instability resulting from deficiency in folate metabolism, initiated by a genetic mechanism, can cause transgenerational outcomes and that environmental and dietary factors can affect DNA methylation intergenerationally in mammals. Several human cohorts like famine and DES exposure demonstrated the intergenerational effect of nutrients and EDCs on disease phenotypes (Bygren et al., 2014; Klip et al., 2002; Veenendaal et al., 2013; 2012). However, whether environmental perturbations induce epigenetic changes and these changes can be transgenerationally inherited through epigenetic mechanisms to influence human disease susceptibility remains unknown. Furthermore, distinguishing between epigenetic inheritance and epigenetic biomarkers remains challenging, especially in humans. Few epidemiological studies that demonstrate intergenerational and possibly transgenerational associations have assessed the following conditions: paternal line food supply, maternal line famine and prenatal nutrition, smoking, and paternal line betel-quid chewing, recently reviewed in (Pembrey et al., 2014).

4.2. Limitations of human epidemiology studies

That we have evidence for transgenerational epigenetic inheritance in animal models with relatively short lifespans and generation times, such as *D. melanogaster* and *C. elegans* is not surprising. In addition to the time required for multi- or transgenerational studies in humans, limited sample availability makes it difficult to study dynamic genome-epigenome interactions (Section 4.4). While often sufficient for identification of stable biomarkers, human samples do not necessarily provide adequate mechanistic insight needed to establish epigenetic inheritance and rule out genetic inheritance, parental effects, and

recapitulation. Furthermore, human samples are often limited to those that are minimally invasive, such as blood, urine, saliva, skin, and post-mortem samples, which are inadequate for studying the complex epigenetic dynamics that occur during development (Section 1.2). Finally, the inability to completely control for other environmental exposures makes it difficult to identify relevant epigenetic and genetic changes for a particular factor and rule out confounding effects from other environmental influences.

4.3. Limitations of interspecies extrapolation of phenotypes

Animal models can be used to study molecular mechanisms in organisms over multiple lifecycles and throughout development. It is tempting to extrapolate phenotypic consequences of environment-induced epigenetic changes (e.g., adverse outcomes from toxicant exposure) observed in animal models to humans. While similar epigenetic mechanisms likely apply to humans, perturbations to these mechanisms may not necessarily affect the same endogenous gene(s) or pathway(s) as in animal models. For example, estimates suggest that mice and humans have 30 percent of imprinted genes in common (Skaar et al., 2012) and IAP elements, responsible for the lability of the agouti epiallele (Sections 2.4 and 3), are not found within the human genome. Therefore, direct interspecies extrapolation of phenotypes identified in rodent models may be insufficient for explaining particular human traits or disease susceptibility. However, other families of transposable elements within the human genome could impart similar characteristics to endogenous human genes. A mechanistic understanding of interspecies genomic and epigenomic differences would be particularly important in extrapolating effects observed in rodents involving imprinted genes, mouse IAPs, and other transposable elements to humans.

One study demonstrated interspecies differences in the regulation of retrotransposons at transcription start sites (TSSs) in mouse and human genomes using multiple cell and tissue types for each species (Faulkner et al., 2009). The comparison used cap analysis gene expression (CAGE) followed by high-throughput sequencing, which provides detailed information on activity at TSSs by capturing and sequencing the first 20 nucleotides from the 5 prime end of full-length cDNAs. They found that approximately 18.1 percent of mouse and 31.4 percent of human TSSs were located within repetitive elements, but only 5.2 percent of mouse and 2.8 percent of human TSSs with more than 100 CAGE tags were from retrotransposons. This finding indicates that although abundant throughout mammalian

genomes, TSSs of retrotransposons tend have lower transcription than non-repeat TSSs. Furthermore, transcription from repetitive element-associated TSSs differed between cell and tissue types. Interestingly, transcription of retrotransposons was poorly conserved between mice and humans, even in comparable tissue types. Although whether this reflects species-specific epigenetic regulation of transposable elements or differences between human and mice retrotransposon families was not determined. Overall, this genome-wide characterization of transcription at repetitive elements in both mouse and human cells/tissues suggests that regulation of retrotransposons in mammalian genomes displays species specificity and that variation also occurs within a species among different developmental stages and tissue types. These differences might need to be taken into account to extrapolate from rodent models to human diseases.

The dependence on transposable elements for both metastable epialleles and inherited paramutations suggests that transposable element-associated genes that escape DNA methylation erasure in both zygotes (Borgel et al., 2010; Smallwood et al., 2011) and PGCs (Guibert et al., 2012; Hackett et al., 2013b) could be candidate regions to explore for heritable epialleles in mammals. With bioinformatic approaches and bisulfite sequencing, a recent study identified retrotransposon-derived candidate epialleles in the mouse genome (Ekram et al., 2012). Identification of similar candidate regions within the rat and human genomes would be useful for understanding interspecies similarities and differences in genomic regions that have the potential to provide insight into potential transgenerational epigenetic inheritance in mammals.

4.4. Genome-epigenome dynamics

DNA methylation exhibits some sequence context specificity, indicating that genetic variation among individuals could result in DNA methylation differences (Schübeler, 2012). At the same time, maintaining genome integrity requires suppression of retrotransposition activity, accomplished via epigenetic regulation (Ishiuchi and Torres-Padilla, 2014). This observation has lead others to propose that environmental factors could result in primary, secondary, or tertiary epimutations, reviewed in (McCarrey, 2014). Briefly, primary epimutations would be initiated by an epigenetic change that is then inherited via epigenetic mechanisms with no changes to DNA sequence (Whitelaw and Whitelaw, 2008). Secondary epimutations would be initiated by a genetic change that causes an epigenetic change that is subsequently

inherited by either genetic or epigenetic mechanisms (Whitelaw and Whitelaw, 2008). Tertiary epimutations would be initiated by an epigenetic change that causes a genetic change(s) (McCarrey, 2012). The apparent reciprocal dependence between the epigenome and genome suggests a dynamic interaction, which makes it difficult to establish whether epigenetic differences are the initial cause or consequence of DNA sequence variation, especially when the only available molecular evidence is static and temporality must be inferred.

5. Potential implications for human environmental health and disease

Without conclusive mechanistic evidence either for or against transgenerational epigenetic inheritance of environment-induced epigenetic changes in mammals, assessing the human health implications is somewhat speculative. Assuming (1) that transgenerational epigenetic inheritance occurs in mammals and we have yet to find a mechanism, (2) that because of their common use in biomedical research we would most likely establish such a mechanism in rodent models, and (3) that environment-induced perturbations to these epigenetic factors could also be inherited, we would still need to extrapolate those mechanisms to human diseases. Molecular mechanisms in the most general sense should be applicable to humans. If, for example, a transposon silencing-based mechanism were to be implicated in any transgenerational disease phenotypes observed in rodent models (Section 2.1), it would be possible for a similar silencing mechanism to occur in humans. However, interspecies differences in families of transposable elements, their insertions sites, and their influences on gene regulation (Levin and Moran, 2011) suggest that similarly regulated genomic features may not affect the same genes in humans as in animal models. Fortunately, the increasing wealth of genomic and epigenomic data should allow us to extrapolate molecular mechanisms from rodent models to identify similar regions of the human genome based on characteristic genomic and epigenomic features. After we identify distinctive features of transgenerational inheritance in the rodent models, our next step would be to uncover details of analogous or homologous mechanisms in humans, and then to identify potential genes affected, and finally, to connect changes in expression of these genes to phenotypic (e.g., disease) consequences.

Alternatively, the observed environment-induced transgenerational phenotypes could be epigenetic manifestations of heritable genetic mutations or transmitted by other non-genetic mechanisms.

Given the rarity of transgenerationally-inherited, environment-induced epigenetic changes in plants, in which mechanisms of transgenerational epigenetic inheritance have been established (Heard and Martienssen, 2014), combined with the rarity of epialleles reported in mammals, other scenarios could explain many examples of transgenerational phenotypes observed in mammals. A recent study assessing DNA methylation of 614 individuals from 117 families consisting of monozygotic or dizygotic twins and their parents found that similarity in DNA methylation between relatives could be accounted for largely by genetic heritability of DNA sequence variants (McRae et al., 2014). Under a genetic inheritance scenario, we would still need detailed mechanistic information to determine affected genes and understand possible human phenotypes, especially if the DNA sequence variations were found to occur in noncoding areas of mammalian genomes.

6. Conclusions

Environmental conditions, such as toxicant exposure and metabolic deficiencies, can influence phenotypes and change epigenetic marks over multiple generations. However, whether these altered epigenetic marks are evidence of transgenerational epigenetic inheritance of environmental effects remains unknown. The demonstrated mechanisms for transgenerational epigenetic inheritance in other eukaryotic organisms and transgenerational epigenetic silencing of transposable elements in the mammalian genome suggests that it would be plausible for similar processes to exist in mammals. Then, it is not difficult to imagine that environmental conditions could alter these epigenetic processes. Although the two epigenetic reprogramming events within the mammalian lifecycle would make the transmission of epigenetic marks across multiple generations unlikely, some repetitive elements and single-copy loci resist germline erasure and some repetitive elements and imprinted loci resist early embryonic erasure. However, the rarity of epialleles and their association with silencing transposable elements in mammals leaves the possibility that toxicant or metabolic-induced epigenetic instability could give rise to *de novo* insertion mutations resulting from reactivation of transposable elements in offspring that were presumed to be isogenic. The mammalian lifecycle further complicates distinguishing between germline transmission and environmental influences, especially parental, as subsequent generations are never truly independent of their predecessors in natural, as opposed to experimental, mammalian systems.

Many transgenerational phenotypes appear to be mediated, if not inherited, by epigenetic processes. As such, distinguishing among epigenetic, genetic, and metabolic contributions to human environmental diseases will continue to be active areas of research and debate. Demonstrating whether transgenerational epigenetic inheritance occurs in mammals will require detailed genetic, epigenetic, transcriptomic, and metabolic profiling of well-controlled animal models with well-documented pedigrees. Novel cell-based and in vitro approaches for studying dynamic molecular processes will be needed to identify epigenetic factors involved and rule out possible confounding from non-germline parental and grandparental influences on offspring, such as those introduced by behavioral, metabolic, or microbiotic factors. Regardless of whether transgenerational phenotypes are transmitted via epigenetic or other mechanisms, extrapolating findings to understand the role that environmental perturbations play in human disease susceptibility necessitates a quantitative and integrated understanding of the interactions among the environment and the genome, epigenome, transcriptome, and metabolome at different stages of human development.

Part B–Aims and Objectives

It is increasingly important to understand how epigenetic states are established and maintained to determine how the plasticity and heritability of epigenetic marks, such as DNA methylation, contribute to mechanisms of toxicity from developmental exposures. Due to the ability of genomic imprints and certain transposable elements to resist either the post-fertilization or germline specification phases of DNA methylation reprogramming(Cowley and Oakey, 2012; Seisenberger et al., 2012), parent-of-origin dependent regulation of imprinted genes and epigenetic silencing of transposable elements are of particular interest for investigating molecular mechanisms of intergenerational and potential transgenerational inheritance from environmental exposures. To identify epigenetic characteristics of these genomic features that would render them vulnerable to epigenetic perturbation, we analyzed whole-genome bisulfite sequencing (WGBS) from mouse embryonic stems cells (mESCs) with disrupted DNA methyltransferase (DNMT) enzymes and from male mice exposed *in utero* to Bisphenol A (BPA).

Mouse embryonic stem cells (mESCs) are pluripotent cells derived from the inner cell mass of preimplantation embryos (Hackett et al., 2013a; Nichols and Smith, 2009). As such, mESCs represent an

early embryonic stage capable of differentiating into either somatic tissues or primordial germ cells (Hackett and Surani, 2012). Therefore, detailed knowledge of how epigenetic marks are maintained throughout the genome at this critical stage of early embryonic development should provide insights into epigenetic inheritance paradigms. Three DNA methyltransferase (DNMT) enzymes are responsible for adding methyl groups to DNA in mammals. DNMT1 is classified as the "maintenance" enzyme because it localizes to replication foci (Leonhardt et al., 1992) and has a preference for hemimethylated DNA (Bestor and Ingram, 1983; Gruenbaum et al., 1982). DNMT3a and DNMT3b are considered *"de novo"* methyltranferases due to their role in establishing new methylation patterns in during embryonic development (Okano et al., 1999). This leads to the traditional two-step model, whereby DNMT3a/3b establishes and DNMT1 maintains DNA methylation patterns (Fig. 1.1). However, this model is limited in its inability to predict the diversity in DNA methylation changes observed throughout the genome in response to environmental or genetic influences.

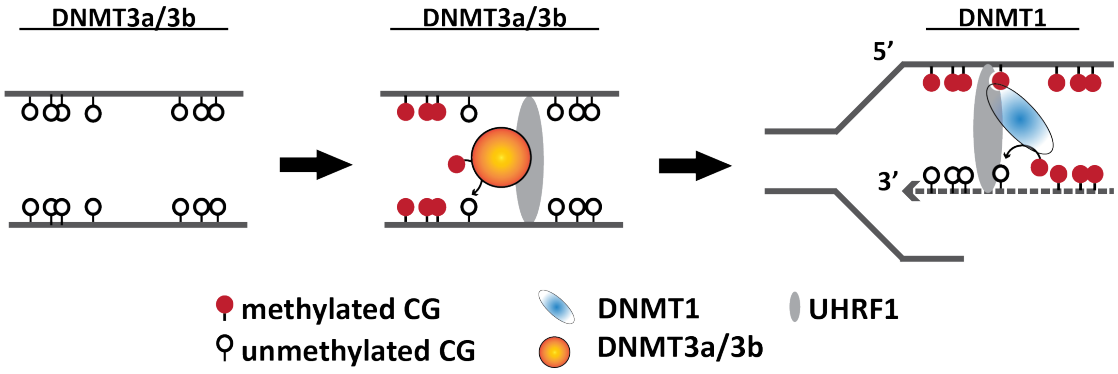


Figure 1.1 Two-step model for DNA methylation.

Previously, we applied whole-genome bisulfite sequencing (WGBS) to characterize methylomes of wildtype (WT) mESCs (J1; ATTC SCRC-1010) and three mutant mESCs with altered DNA methyltransferase (DNMT) enzymes: loss of "maintenance" DNMT (1KO; *Dnmt1*$^{-/-}$), loss of "*de novo*" DNMTs (DKO; *Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$), and loss of "maintenance" and "*de novo*" DNMTs (TKO; *Dnmt1*$^{-/-}$/*Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$). We demonstrated that, in general, CpG sites throughout the mouse genome rely on activity from both "maintenance" and "*de novo*" DNMTs in mESCs to some extent (Li et al., 2015). We also identified loci that were predominantly DNMT1 maintained or preferentially DNMT3a/3b maintained. DNMT1-maintained regions were enriched in long terminal repeats (LTRs), including Intracisternal A

Particle (IAP) elements. DNMT3a/3b-maintained regions were enriched in long interspersed nuclear elements (LINEs). This lead to our proposed "coordination and division of labor" model for DNA methylation maintenance (Fig. 1.2). Because of this, we postulated that DNA methylation maintenance paradigms of genomic regions might influence their susceptibility to heritable toxicant-induced epigenetic perturbation.

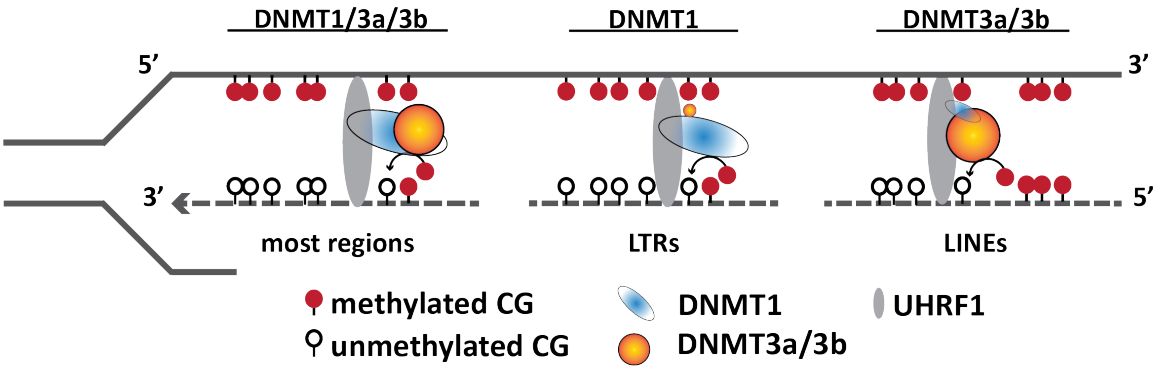


Figure 1.2 Coordination and division of labor model for DNA methylation maintenance.

Aim 1: Identify genomic regions in mESCs that share DNA methylation characteristics found at germline-imprinted regions

Genomic imprints are established in a sex-specific manner during germ cell development, leading to germline differentially methylated regions (gDMRs) between sperm and oocytes (MacDonald and Mann, 2013). Unlike the majority CpG sites throughout the genome, DNA methylation at imprinted gDMRs is not erased during post-fertilization reprogramming (Smith et al., 2014). This unique paradigm of epigenetic inheritance results in one highly methylated allele and one undermethylated allele in diploid mammalian zygotes. Furthermore, loss of DNA methylation in *Dnmt1*$^{-/-}$ mESCs can be partially rescued at most genomic regions with exogenous *Dnmt1* cDNA expression (Tucker et al., 1996). In contrast, loss of DNA methylation at several imprinted differentially methylated regions (DMRs) could not be rescued (Tucker et al., 1996). Chapter 2 focuses on developing and applying bioinformatic methods to identify regions that share the characteristics of bimodal methylation patterns and non-restorable DNMT1-dependent DNA methylation loss throughout the mouse genome.

Aim 2: Determine whether DNA methylation characteristics associated with gDMRs and transposable elements increase susceptibility to epigenetic perturbation from prenatal exposure to Bisphenol A (BPA)

Exposure to endocrine disrupting chemicals, such as BPA, during early development can increase later life disease risk for cancers (Acevedo et al., 2013; Bindhumol et al., 2003; Dhimolea et al., 2014; Fenton et al., 2012; Ho, 2006; Jenkins et al., 2009; Maffini et al., 2006; Prins and Ho, 2010; Soto et al., 2013; Tang et al., 2012; Weinhouse et al., 2014), cardiovascular disease (Lang et al., 2008; Melzer et al., 2010), metabolic abnormalities (Angle et al., 2013; Chou et al., 2011; Harley et al., 2013; Hatch et al., 2010; Li et al., 2014; Miyawaki et al., 2007; Somm et al., 2009; Stojanoska et al., 2017; Wei et al., 2011; Zoeller et al., 2005), and neurological disorders (Nakagami et al., 2009; Perera and Herbstman, 2011; Zhou et al., 2009). In mice, perinatal exposure to BPA alters DNA methylation at several imprinted genes (Susiarjo et al., 2013) and retrotransposon-associated metastable epialleles (Anderson et al., 2012; Dolinoy et al., 2007b). In Chapter 3, we examine whether genomic regions sharing epigenetic characteristics observed in mESCs at imprinted loci or transposable elements render them more likely to experience BPA-induced DNA methylation changes. Identification of regions of high susceptibility to epigenetic toxicity should improve the ability of researchers to connect molecular targets with disease phenotypes to develop a mechanistic understanding of how developmental exposures lead to diseases later in life and across multiple generations.

Aim 3: Determine potential functional relevance of epigenetic perturbation for metabolic phenotypes

Maternal gestational exposure to BPA has been demonstrated to lead to metabolic abnormalities in male offspring for two generations, in mice (Susiarjo et al., 2015). To gain a broad perspective into potential DNA methylation-mediated effects, in Chapter 4, we use whole-genome bisulfite sequencing to identify pathways with altered DNA methylation in fetal livers of male mice that were exposed *in utero* to BPA.

Chapter 2

Two approaches reveal a new paradigm of 'switchable or genetics-influenced allele-specific DNA methylation (ASM)' with potential in human disease

(*SNM developed algorithms for NORED and MethylMosaic, implemented bioinformatics analyses, and wrote Results and Methods Sections pertaining to those analyses. SNM contributed to writing Abstract, Introduction, and Discussion.*)

Abstract

Imprinted genes, important for development and neurological function, are vulnerable to nutritional/environmental influences during early embryonic development, thereby contributing to the onset of disease in adulthood. Allelic methylation at several germline imprints has been reported as DNMT1-dependent. However, which of two epigenetic attributes, DNMT1-dependence or allelic methylation, renders them susceptible to environmental stressors has not been determined. We therefore sought to detect regions throughout the genome that exhibited DNMT1-dependence and/or had bimodal methylation patterns. Herein we have developed an approach, referred to as NORED, to identify 2468 maintenance DNMT1-dependent DNA methylation patterns. We further use a genetic variation-independent approach (referred to as MethylMosaic) to detect 2487 regions with bimodal distribution of hypermethylated or hypomethylated reads. Combined, two approaches identified 207 regions, including known imprinted germline ASMs, which were lost in DNMT-deficient cells and had bimodal methylation. Unexpectedly, we identified regions that although both DNMT1-dependent and bimodal, do not display parent-of-origin dependent allelic methylation in reciprocal cross ESCs. Further focused studies on candidate ASMs from two loci, *Hcn2* (epilepsy and chronic pain) and *Park7* (Parkinson's disease), revealed a new paradigm. Genetic variation in Cast allele at *Hcn2* locus introduces a transcription factor binding site for MTF-1 that coincides with a predisposition for allelic hypomethylation in a reciprocal cross with either C57 or 129 strains, whereas genetic differences at *Park7* locus do not affect allelic ratio of highly and lowly methylated reads. That is, each allele of the latter exhibits similar propensity to be either hypo- or hypermethylated, suggesting a "random, switchable" ASM. Together with published results, our data on ASMs raise a possibility of regional "autosomal chromosome inactivation (ACI)"-like mechanism controlling a subset of autosomal genes. Collectively, our investigations open a new window to understand pathogenesis of many common diseases.

1. Introduction

Environmental factors impact human health (Dai and Wang, 2014; Jirtle and Skinner, 2007; Rappaport and Smith, 2010). For example, a high fat diet contributes to the pathogenesis of obesity and type 2 diabetes (Heydemann, 2016; Leung et al., 2016), and low nutrition during early embryonic development may link to cardiovascular diseases in later life (Barker and Osmond, 1986); and exposure to endocrine disrupting chemicals such as Bisphenol A increases the susceptibility to diseases including neurological disorders (Nakagami et al., 2009; Perera and Herbstman, 2011; Zhou et al., 2009), heart diseases (Lang et al., 2008; Melzer et al., 2010) and cancers (Acevedo et al., 2013; Bindhumol et al., 2003; Dhimolea et al., 2014; Fenton et al., 2012; Ho, 2006; Jenkins et al., 2009; Maffini et al., 2006; Prins and Ho, 2010; Soto et al., 2013; Tang et al., 2012; Weinhouse et al., 2014). These factors contribute to disease pathogenesis, largely via epigenetic mechanisms (Choi and Friso, 2010; Dai and Wang, 2014; Feil and Fraga, 2012; Jirtle and Skinner, 2007; Leung et al., 2016). Altered epigenetic patterns have been involved in numerous human diseases. A notable example is that cancer cells are characterized by genome-wide hypomethylation and region-specific hypermethylation. Additional examples include abnormal hypomethylation of genes involved in autoimmune diseases such as systemic lupus erythematosus (Jeffries and Sawalha, 2011). For mechanistic insights of such alteration in diseases, investigators have been focusing on exposure-induced changes in expression or function of DNA methyltransferases (DNMTs) (Choi and Friso, 2010). Many publications report altered expression of DNMT1, the maintenance enzyme for transmitting DNA methylation patterns in generations of somatic cells (Choi and Friso, 2010). However, DNMT-dependent alterations of DNA methylation in diseases remain to be determined. A systematic examination of DNMT-dependent methylation regions via global profiling of DNMT-deficient cells will significantly improve our understanding of such alterations in diseases.

Imprinted genes are associated with monoallelic methylation and monoallelic gene expression. As such, they are considered particularly vulnerable to environmental exposure (Lambertini, 2014; Murphy and Hoyo, 2013; Susiarjo et al., 2013). Monoallelic expression of imprinted genes is controlled by germline allele-specific methylation (ASM; traditionally called as differentially-methylated-region (DMR)). For these germline ASMs (gASMs), CpG sites at the imprinting control regions of one parental allele are

methylated, whereas these of another allele are unmethylated. When examined by bisulfite Sanger sequencing, half of PCR clones in those regions are hypermethylated and half are hypomethylated, thereby showing bimodal methylation patterns. Importantly, these gASMs/DMRs are considered stable during the cycles of global demethylation and remethylation during early embryo development (Bartolomei and Ferguson-Smith, 2011). Once these ASMs have been altered due to exposure in sperms and/or oocytes, such alteration could be carried as "epigenetic memories" to somatic cells. Numerous studies focus on the alteration of these ASM-controlled imprinted genes for understanding of "developmental origin of adult disease" and "transgenerational epigenetic inheritance" of disease (Martos et al., 2015; Murphy and Hoyo, 2013). However, most current studies use a candidate approach that targets several imprinted ASMs (or imprinted genes), particularly *Igf2* (Murphy and Hoyo, 2013). For example, in a widely cited study, the methylation level of several CpG sites of *Igf2* ASM from Dutch famine patients was shown to be maintained after decades (Heijmans et al., 2008). Limited work has been done to systematically examine all gASMs, including their exact sizes and the possibility of unknown gASMs. Such a genome-wide examination will be helpful for understanding the pathogenesis and diagnosis of human diseases/syndromes.

In addition to the above-mentioned parent-of-origin specific monoallelic expression, the mammalian genome has a surprisingly large number of genes showing random monoallelic expression (RME) (Deng et al., 2014; Eckersley-Maslin et al., 2014; Gendrel et al., 2014; Lo et al., 2003; Yan et al., 2002). Although genetic variants can affect expression, such monoallelic expression is unexpected, as the conventional notion is that non-imprinted genes on autosomal chromosomes should be either bialellically expressed or biallelically repressed. Monoallelic gene expression was previously thought to occur only at imprinted loci or genes from X chromosomes for which one chromosome (paternal or maternal) is randomly inactivated in females (Tycko, 2010). However, recent studies found that the monoallelic expression of non-imprinted, autosomal genes does not seem to be a sporadic phenomenon, but a conserved feature in both the mouse genome (Deng et al., 2014; Eckersley-Maslin et al., 2014; Gendrel et al., 2014) and the human genome (Lo et al., 2003; Yan et al., 2002). While RMEs seem to occur frequently, the underlying mechanism remains elusive. Two recent investigations suggest epigenetic mechanism cannot account for RME (Eckersley-Maslin et al., 2014; Gendrel et al., 2014). In addition, the role of

monoallelically expressed genes in both development and diseases such as cancers (Chen et al., 2008) needs urgent investigation (see the significance as exemplified by this work later).

Monoallelic methylation (i.e., imprinted ASM) or expression has been provided as a rationale to explain why imprinted genes are susceptible to nutritional and environmental influences. With similar monoallelic methylation/expression, it is reasonable to expect that other ASMs or RMEs would demonstrate similar vulnerability. However, the molecular explanation for these expectations remains to be fully explored. Early studies have identified several gASMs as highly DNMT1-dependent in preimplantation embryos (Tucker et al., 1996). Specifically, global methylation substantially recovered in 'rescued' DNMT1-deficient embryonic stems cells (ESCs), whereas the methylation at several gASMs, which had been abolished in DNMT1-deficient ESCs, was not restored in 'rescued' DNMT1-deficient ESCs (Tucker et al., 1996). Notably, while DNMT1-deficiency is embryonic lethal (Li et al., 1992), 'rescued' DNMT1-deficient ESCs could contribute to viable adult chimeras (Tucker et al., 1996). In contrast, overexpression of DNMT1, which resulted in *de novo* methylation at the unmethylated allele of *Igf2*, was embryonic lethal (Biniszkiewicz et al., 2002). Taken together, these early studies suggest that while DNMT1 is required to maintain imprinted ASMs, hypermethylation is not compatible with embryonic viability, whereas hypomethylation of the methylated allele for some gASMs may be tolerated. If the finding of DNMT1-dependent DNA methylation loss in preimplantation embryos can be extended to additional imprinted gASMs, then this apparent vulnerability could provide an alternate explanation for why gASMs of imprinted genes are considered especially susceptible to environmental influences. Yet, this raises the question as to which of two epigenetic attributes, DNMT1-dependence or allelic methylation, renders gASMs susceptible to environmental stressors. Furthermore, it raises the question as to whether additional genomic regions display non-restorable DNMT1-dependent methylation loss.

To provide mechanistic insights for questions above, herein we investigate regions throughout the genome that exhibit DNMT1-dependence and/or have bimodal methylation patterns. We begin with examining DNA methylation patterns associated with genomic imprinting, which is complementary to our previous investigation of methylation patterns associated with gene transcription and genome stability (i.e., suppression of transposable elements) (Li et al., 2015). Based on the loss of gASMs/DMRs in *Dnmt1*

knockout (1KO) ESCs and failure to restore the loss in 1KO cells with exogenous expression of *Dnmt1* cDNA, we developed a new approach, "non-rescued DMR (NORED)", identifying genomic regions dependent on DNMT1. Among these regions, many are *bona fide* imprinted gASMs with the expected bimodal methylation patterns, as unveiled by our "MethylMosaic" analyses. In addition to the known imprinted gASMs, MethylMosaic analyses also identify genomic regions showing bimodal methylation patterns. We next generate four independent mouse ESC lines from hybrid mice and demonstrate that some NORED regions with bimodal methylation patterns show allelic methylation, but in a parent-of-origin independent fashion. Intriguingly, genetic differences at *Hcn2/Polrmt* locus predisposed Cast allele to be hypomethylated in cross with either 129 or C57, whereas genetic differences at *Park7* locus did not. Their shared features (e.g., allelic hyper-/hypomethylation) with X chromosome inactivation (XCI) raise the possibility that many genes on autosomal chromosomes are controlled by regional "autosomal chromosome inactivation (ACI)."

2. Results

All gASMs are lost in DNMT1-deficient ESCs, whereas specific loci exhibit resistance to methylation loss in DNMT3a/3b-deficient ESCs

To determine DNA methylation patterns' impacts on gene transcription, genome stability, and genomic imprinting, we have characterized the base resolution DNA methylomes of wild type (WT, J1 ESC line) and DNMT-deficient ESCs, including the loss of maintenance DNMT1 (*Dnmt1*$^{-/-}$, 1KO), of two *de novo* DNMT3a/3b (*Dnmt3a*$^{-/-}$/*3b*$^{-/-}$, DKO), and of all three (*Dnmt1*$^{-/-}$/*3a*$^{-/-}$/*3b*$^{-/-}$, TKO; Supplemental Fig. S2.1A–B, D–E). Genome-wide, average methylation was 0.727, 0.176, 0.157, and 0.006 in WT, 1KO, DKO, and TKO, respectively. Previously, we reported on the distinct roles of DNMT1-dependent and -independent methylation patterns in suppression of transposable elements and the complete hypomethylation on induction of only small number of genes in the mouse genome (Li et al., 2015). Herein we focused our study on DNA methylation patterns at imprinted loci.

DNA methylation is essential for genomic imprinting (Branco et al., 2008; Li et al., 1993). To determine the extent to which gASMs are dependent on maintenance DNMT1 or *de novo* DNMT3a/b for

methylation maintenance, we compared the methylation levels of 21 well-characterized gASMs found in common between two sources (Arnaud, 2010; MacDonald and Mann, 2013) among WT, 1KO, DKO, and TKO ESCs (Fig. 2.1A and Supplemental Table S2.1). Except for 1KO versus DKO, pairwise comparisons were significant (Bonferroni corrected $p < 0.05$). Methylation levels of imprinted gASMs were significantly reduced in 1KO, DKO, and TKO compared to WT (Bonferroni corrected $p = 1.1 \times 10^{-5}$; Fig. 2.1A and Supplemental Table S2.1). Compared to TKO, imprinted gASMs had significantly higher methylation in both 1KO (Bonferroni-corrected $p = 5.7 \times 10^{-6}$) and DKO (Bonferroni-corrected $p = 6.3 \times 10^{-4}$; Fig. 2.1A and Supplemental Table S2.1). However, methylation of imprinted gASMs did not differ significantly between 1KO and DKO (Bonferroni-corrected $p = 0.085$; Fig. 2.1A and Supplemental Table S2.1). Therefore, in general, methylation maintenance at known gASMs is dependent on activity from DNMT1 and DNMT3a/3b in ESCs.

Targeted studies at a few gASMs have reported "near complete" abolishment of allelic methylation upon the loss of DNMT1 in preimplantation embryos (Biniszkiewicz et al., 2002; Hirasawa et al., 2008). To compare these reported results with our findings, we examined the methylation level at individual loci. Although "near complete" methylation loss was not quantitatively defined (Hirasawa et al., 2008), loci presented had at most five percent methylation. From our data of 1KO ESCs, only two gASMs had greater than five percent methylation: *Rasgrf1* (12.2%) and *H19* (6.6%; Fig. 2.1A and Supplemental Table S2.1). Methylation at other imprinted loci, including *Gtl2* (*Meg3*) and *Mest*, is indeed abolished (Fig. 2.1A–C and Supplemental Table S2.1). This is consistent with reports that DNMT1 is necessary for the maintenance of imprinted gASMs (Biniszkiewicz et al., 2002; Hirasawa et al., 2008).

Contrary to the indispensable role of maintenance DNMT1 above, *de novo* activities of DNMT3a/3b were reported to be dispensable for the maintenance of two paternally methylated DMRs, *H19* and *Gtl2*, as neither DMR was affected. For another paternally methylated locus, *Rasgrf1*, the reduced methylation was attributed to an unusual repeat structure at this region (Hirasawa et al., 2008). However, using our base-resolution methylome data (which covers the entire gASM instead of a portion as in (Hirasawa et al., 2008)), we found that methylation at *H19* and *Rasgrf1* was substantially reduced, whereas *Gtl2* only decreased by 0.18 (Fig. 2.1A–B and Supplemental Table S2.1). In total, eight gASMs (which

represent both maternally and paternally methylated loci) had greater than five percent methylation in DKO (DNMT3a/3b-deficient and DNMT1 intact) ESCs: *Gtl2* (30.6%), *Peg13*/*Trappc9* (16.2%), *Rasgrf1* (15.9%), *Peg3* (15.0%), *Inpp5f-v2* (12.7%), *H19* (12.3%), *Plagl1* (9.5%), and *Nespas-GnasXL* (8.1%; Supplemental Table S2.1). Strikingly, *Mest* and remaining gASMs were completely abolished (Fig. 2.1A, C and Supplemental Table S2.1), suggesting an indispensable role of DNMT3a/3b in the maintenance of certain gASMs.

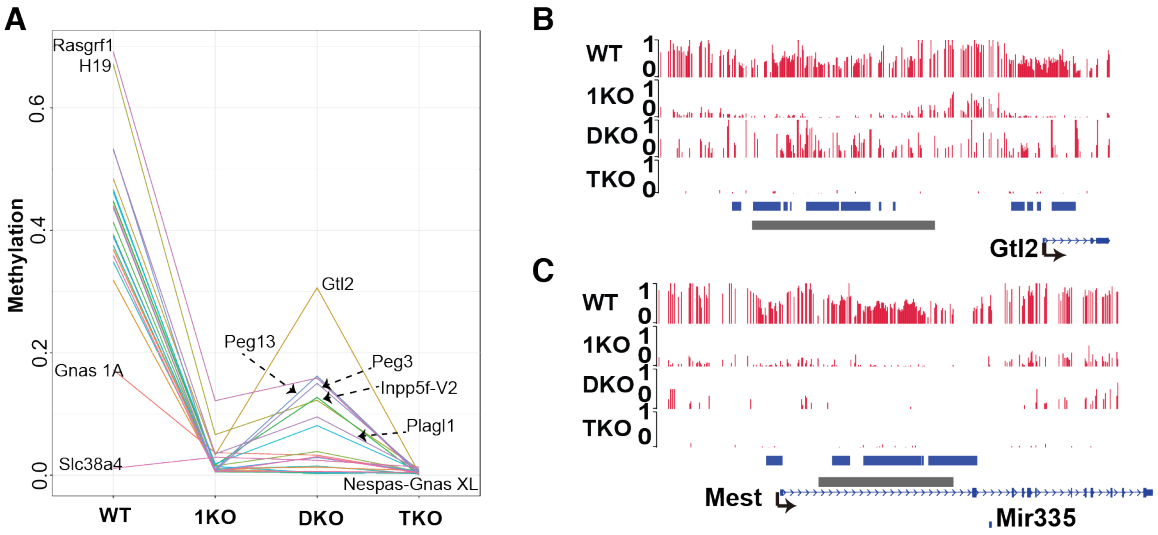


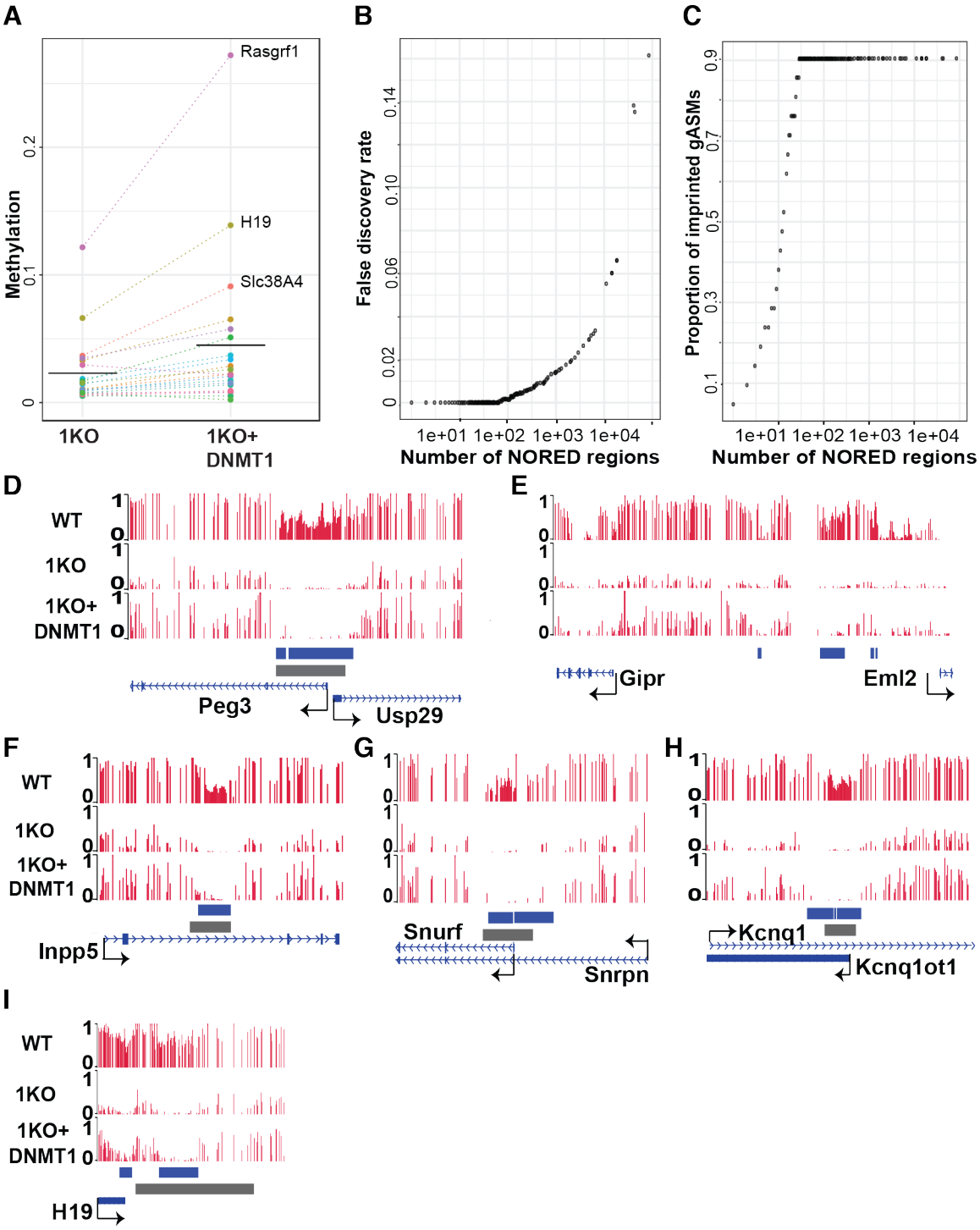
Figure 2.1 Germline ASMs (gASMs) are lost in DNMT1-deficient ESCs, whereas specific loci exhibit resistance to methylation loss in DNMT3a/3b-deficient ESCs. (A) Genome-wide profiling of gASM methylation level in WT and DNMT mutant mouse ESCs. Each line represents a single gASM. Eleven gASMs (unlabeled) have average methylation levels consistent with expectation for one methylated allele and one methylated allele in WT, experience near complete loss of methylation in 1KO and DKO, and complete methylation loss in TKO. *Rasgrf1* and *H19* gASMs have methylation levels somewhat higher than expectation for one methylated allele and one methylated allele in WT, retain partial methylation in 1KO and DKO, and experience complete loss of methylation in TKO. Six gASMs (*Peg13, Gtl2, Peg3, Inpp5f*-V2, *Plagl1, and Nespas-Gnas* XL) have methylation levels consistent with expectation for one methylated allele and one methylated allele in WT, experience near complete loss of methylation in 1KO, retain partial methylation in in DKO, and experience complete loss of methylation in TKO. *Gnas* 1A has partial methylation in WT and experiences near complete loss of methylation in WT, 1KO, and TKO. *Slc38a4* has low methylation in in WT, 1KO, DKO, and TKO. Genome-wide, average methylation was 0.727, 0.176, 0.157, and 0.006 in WT, 1KO, DKO, and TKO, respectively. See Supplemental Table S2.1 for locus-specific methylation levels. **(B)** *Gtl2* gASM is abolished in 1KO cells. In contrast, DKO cells retain partial methylation patterns. **(C)** Methylation at *Mest* locus is completely abolished in both 1KO and DKO cells. **(B–C)** Blue bars indicate NORED regions. Gray bars indicate gASMs.

Altogether, we conclude that DNMT1 is necessary, but not sufficient to maintain methylation at gASMs and specific loci exhibit partial resistance to methylation loss in the absence of DNMT3a/3b. These facts reveal the previously unappreciated coordination between *de novo* and maintenance activities to maintain methylation patterns. These facts are different from the traditional view/model of 'DNMT3a/3b for initiating methylation and DNMT1 for maintaining afterward.'

Loss of methylation was not rescued at gASMs

Recognizing that important gASMs require DNMT1 for maintenance of DNA methylation, we wondered whether loss of methylation at these loci could be restored once it had been abolished. Literature search indicated that exogenous expression of *Dnmt1* cDNA did not restore methylation at a few gASMs (Biniszkiewicz et al., 2002). To extend this finding to other well-characterized gASMs, we expressed *Dnmt1* cDNA in 1KO cells to characterize the base resolution DNA methylome of 'rescued 1KO' ESCs (*Dnmt1*$^{-/-}$ + *Dnmt1* cDNA, r1KO; Supplemental Fig. S2.1C). Global average methylation in r1KO (0.369) increased to approximately 50.8 percent of WT levels, whereas average 1KO levels were 24.1 percent that of WT (Supplemental Fig. S2.1F). Therefore, we consider global methylation to be substantially restored in r1KO ESCs. In contrast, average methylation for gASMs decreased to 0.023 (5.4 percent of WT levels at these loci) in 1KO and recovered to only 0.045 (10.8 percent of WT levels at these loci) in r1KO (Fig. 2.2A and Supplemental Table S2.1).

Figure 2.2 Loss of methylation was not rescued at gASMs and other specific loci. (A) Profiling of gASM methylation level in 1KO and DNMT1-rescued 1KO (r1KO) ESCs. For most gASMs, low methylation in 1KO is not substantially increased by exogenous expression *Dnmt1* cDNA in r1KO ESCs. Genome-wide, average methylation was 0.369 for r1KO cells. See Supplemental Table S2.3 for locus-specific details. **(B)** False discovery rate of NORED. **(C)** The 19 gASMs rank among top 29 NORED regions. **(D)** Deficiency of methylated CpG sites (red bars) in 1KO and DNMT1-rescued 1KO cells can be used to identify and demarcate the known *Peg3* DMR. **(E)** NORED identifies *Gipr*/*Eml2* locus (imprinted status unknown). **(F)** NORED identifies, gASM in *Inpp5f* locus. **(G)** NORED improves the demarcation of gASMs for Prader-Willi and Angelman syndromes (*Snrpn*/*Snurf*) **(H)** NORED demarcates larger region at *Kcnq1/Kcnqot1*. **(I)** NORED successfully identifies gASM and somatic ASM at the *H19/Igf2* locus. **(D–I)** Blue bars indicate NORED regions. Gray bars indicate gASMs.


A
Rasgrf1
H19
Slc38A4
Methylation
0
0.1
0.2
1KO
1KO+
DNMT1
B
False discovery rate
0
0.02
0.06
0.10
0.14
1e+01 1e+02 1e+03 1e+04
Number of NORED regions
C
Proportion of imprinted gASMs
0.1
0.3
0.5
0.7
0.9
1e+01 1e+02 1e+03 1e+04
Number of NORED regions
D
WT
1KO
1KO+
DNMT1
Peg3
Usp29
E
Gipr
Eml2
F
WT
1KO
1KO+
DNMT1
Inpp5
G
Snurf
Snrpn
H
Kcnq1
Kcnq1ot1
I
WT
1KO
1KO+
DNMT1
H19

NORED: genome-wide detection of DNMT1-dependent methylation that is not recovered once abolished

We took advantage of the previous observation to develop NORED, a new method to systematically identify genomic regions with non-rescued DMR. These NORED regions must have sufficient methylation in WT, near complete loss of methylation in 1KO, and minimum recovery of methylation in r1KO. We performed a receiver operating characteristic (ROC)-like analysis using a permutation-based approach to estimate the false positives at various methylation cutoffs for WT, 1KO, and r1KO ESCs (Supplemental Fig. S2.2A). At a false positive rate (FPR) of 0.01, 70 percent of CpG sites within gASMs had at least 25 percent methylation in WT, at most 5 percent methylation in 1KO, and at most 12.5 percent methylation in r1KO (Supplemental Fig. S2.2A). By contrast, genome-wide only 3.3 percent of CpG sites met these criteria. We then clustered individual CpG sites into regions and ranked the resulting NORED regions based on the number of CpG sites included and the proportion of consecutive CpG sites that met the criteria (see Methods). To determine the false discovery rate (FDR) for NORED, we applied the clustering and scoring algorithms to permuted data and estimated the average FDR based on twenty permutations.

NORED analyses identified 2468 regions at FDR = 0.02 (Fig. 2.2B, Supplemental Fig. S2.2B, and Supplemental Table S2.2). The highest ranked 207 regions (FDR < 5×10^{-3}) are presented in Table 2.1. Only two gASMs, *Slc38a4* and *Gnas 1A*, were not identified by NORED because neither of these regions had sufficient methylation in WT (Fig. 2.1A). Strikingly, all remaining 19 established gASMs had at least one NORED region within the highest 29 ranked regions (FDR < 6.8×10^{-4}; Fig. 2.2B–C and Supplemental Fig. S2.2B).

As exemplified by gASMs of *Peg3*, *Inpp5f*, *Snrpn*/*Snurf*, *Kcnq1ot1*/*Kcnq1*, and *H19* from chromosome 7, NORED identified well-established gASMs (Fig. 2.2D, F–I) and potential gASMs (MacDonald and Mann, 2013) at *Cdh15* (Chr 8qE1) and *Nnat/Blcap* (Chr 2qH1) (Table 2.1). At *Peg3*, NORED detected two regions overlapping the known gASM (gray bar in Fig. 2.2D): one 646 bp region that aligns with the established gASM start site and a second 4.2 kb region that covers the majority of the 4.5 kb gASM, but extends beyond the end site (Fig. 2.2D). At *Inpp5*, a single 1.1 kb NORED region covers the majority of the 1.4 kb gASM (Fig. 2.2F). Our NORED analyses defined two tandem ASMs of the

Snurf/Snrpn locus as 2.1 kb and 3.3 kb, respectively (Fig. 2.2G). Three NORED regions of 1.7 kb, 116 bp, and 1.6 kb were detected at *Kcnq1ot1*/*Kcnq1*, extending beyond the 2.1 kb gASM (Fig. 2.2H). *H19* had a 2.4 kb NORED region within the larger reported gASM (7.3 kb; Fig. 2.2I). NORED regions were also detected for *Rasgrf1:* one small (104 bp) within the 8.0 kb gASM and one larger 2.0 kb region extending beyond the end site of the gASM (Table 2.1 and Supplemental Table S2.2).

Additional regions near, but not overlapping with known gASMs were discovered at *Mest* and *Gtl2* (*Meg3*) (Fig. 2.1B–C, Table 2.1). NORED also identified additional imprinted ASM that are not considered gASM (Thorvaldsen et al., 2002), such as somatic ASM at *H19* promoter (774 bp; Fig. 2.2I). Furthermore, we identified a 1.8 kb NORED region near *Gab1*, which is reported to have imprinted gene expression (Babak et al., 2015); however, no gASM close to this region on Chr 8qC2 has been reported (Table 2.1). Finally, NORED identified regions with unknown imprinting or ASM status, as exemplified by the 178 bp, 1.1 kb, 170 bp, and 91 bp intergenic regions between *Gipr* and *Eml2* (Fig. 2.2E). This locus is within Chr 7qA3 and the nearest known gASM (Peg3/Usp29) is upstream within Chr 7qA1; the nearest downstream known gASM (Snrpn/Snurf) is within a Chr 7qB5. The boundaries of known ASMs and other NORED regions are presented in Fig. 2.2, Table 2.1, and Supplemental Table S2.2.

Table 2.1 Identification, demarcation, and characterization of highest ranked NORED

Chromosome location	Nearest gene(s)
chr1:33162585-33162912	*Khdrbs2/Prim2*
chr1:63199813-63200708	***Gpr1/Zdbf2****
chr1:93825020-93825457	*D2hgdh*
chr1:93825898-93826406	*D2hgdh*
chr1:131148330-131149072	*Eif2d/Dyrk3**
chr1:134328756-134329228	*Ppfia4*
chr2:28370200-28370811	*Ppp1r26/Olfm1*
chr2:31304679-31305167	*Ncs1/Ass1/*Gm5424
chr2:32630974-32631585	*Ak1*
chr2:34671610-34671792	*Gapvd1/Mapkap1*
chr2:39806681-39806881	*Ppp6c/Lrp1b*
chr2:59348096-59348285	5330411J11Rik
chr2:105511256-105512346	*Rcn1/Pax6os1*
chr2:130365512-130365714	*Cpxm1*
chr2:131044421-131045271	*Gfra4/Adam33*
chr2:152686347-152686904	***Mcts2/H13****
chr2:152686927-152687288	***Mcts2/H13****
chr2:157559602-157560417	*Nnat/Blcap**
chr2:157561077-157561984	*Nnat/Blcap**
chr2:158614669-158614805	*Slc32a1*
chr2:162485692-162485820	*Ptprt*
chr2:166522472-166522714	*Prex1/*5031425F14Rik
chr2:174293692-174298026	***Nespas/Gnas****
chr2:174299016-174300450	***Nespas/Gnas****
chr2:174298047-174298914	***Nespas/Gnas****
chr2:174300469-174300620	***Nespas/Gnas****
chr2:181307114-181307303	*Stmn3*
chr3:3194456-3199479	*Hnf4g/Cr2*
chr3:94413377-94413565	*Tdrkh*
chr4:43629280-43629987	*Npr2/Rgp1*
chr4:43992511-43993691	*Ccin/Clta*
chr4:93044119-93048203	*Tusc1/Izumo3*
chr4:121052686-121053078	*Col9a2*
chr4:136225076-136225355	*Asap3**
chr4:138677545-138677765	*Ubxn10/Vwa5b1*
chr4:139142929-139143492	*Minos1/*Gm16287
chr4:140814187-140814664	*Gm13032/Padi1*
chr4:140869176-140869307	4930515B02Rik*/Padi1*
chr4:145514859-145516993	Gm13212*
chr4:147808990-147809826	Gm13157*
chr4:150039937-150040661	*Mir34a/H6pd*
chr4:150879968-150880071	*Errfi1/Park7**
chr5:3733077-3733199	*Ankib1*

* Indicates that NORED region overlaps bimodal (MethylMosaic) region
Bold indicates NORED region overlaps known imprinted germline DMR

Table 2.1 (continued) Identification, demarcation, and characterization of highest ranked NORED

Chromosome location	Nearest gene(s)
chr5:35313365-35313892	*Adra2c*/4930478P22Rik*
chr5:36830677-36831279	*Man2b2/Ppp2r2c*
chr5:37040212-37041347	*Jakmip1/Wfs1*
chr5:52515885-52516367	*Ccdc149/Lgi2*
chr5:105732003-105732105	*Lrrc8d*
chr5:105732186-105732404	*Lrrc8d*
chr5:116082856-116083275	*Tmem233*
chr5:117348264-117349083	*Wsb2/Vsig10**
chr5:118268910-118269285	2410131K14Rik*/Med13*
chr5:120761506-120761888	*Oas3*
chr5:120783028-120783454	*Oas3/Oas1e*
chr5:136245578-136246023	*Sh2b2/Cux1*
chr5:137071619-137072273	*Serpine1**
chr5:143128699-143129529	*Rnf216/Rbakdn*
chr5:143133061-143133869	*Rnf216/Rbakdn*
chr6:4746229-4747314	***Sgce****
chr6:4747440-4747696	***Peg10****
chr6:4747780-4747996	***Peg10****
chr6:4748054-4749480	***Peg10****
chr6:30732981-30733591	*Mest/Cep41**
chr6:30735435-30736116	***Mest****
chr6:30736607-30738780	***Mest****
chr6:30739031-30740864	***Mest/Mir335****
chr6:34948776-34948940	*Stra8*/2010107G12Rik
chr6:37464011-37464558	*Creb3l2/Akr1d1*
chr6:39269183-39269977	*Kdm7a/Slc37a3*
chr6:58906186-58907095	***Nap1l5/Herc3****
chr6:85378234-85378818	*Rab11fip5/Noto*
chr6:107531195-107531557	*Lrrn1/Inpp5f*
chr6:115729908-115730603	*Tmem40 **
chr6:119594548-119594870	*Wnt5b*
chr6:125349316-125349717	*Tnfrsf1a/Scnn1a**
chr6:130973317-130980712	*Klra2/Klra22/Klra15*
chr6:136518731-136518889	*Atf7ip*
chr7:4531968-4532355	*Dnaaf3*
chr7:6727076-6727722	***Peg3****
chr7:6727892-6732060	***Peg3****
chr7:10324647-10325420	*Vmn1r66/Vmn1r67*
chr7:16893755-16894739	*Gng8*
chr7:19175647-19176796	*Eml2/Gipr**
chr7:19811098-19811297	*Bcl3**
chr7:24611394-24611596	*Phldb3*

* Indicates that NORED region overlaps bimodal (MethylMosaic) region
Bold indicates NORED region overlaps known imprinted germline DMR

Table 2.1 (continued) Identification, demarcation, and characterization of highest ranked NORED

Chromosome location	Nearest gene(s)
chr7:25301189-25301458	*Prr19*
chr7:25903349-25903407	*Cyp2b10**
chr7:29211649-29211960	*Catsperg1*
chr7:33836946-33837410	*Scgb1b24/Scgb2b26*
chr7:41964212-41974972	*Vmn2r59/Vmn2r58*
chr7:45340239-45340780	*Ppfia3*
chr7:60002938-60005042	***Snurf****
chr7:60005123-60008428	***Snurf****
chr7:102289205-102289935	*Stim1*
chr7:128687633-128688728	***Inpp5f****
chr7:142577780-142578554	*H19**
chr7:142580201-142582623	***H19/IGF2****
chr7:143293662-143295410	***Kcnq1ot1/Kcnq1****
chr7:143295635-143297239	***Kcnq1ot1/Kcnq1****
chr8:3656464-3657513	*Retn*
chr8:27222728-27224119	*Adrb3/Got1l1**
chr8:56623073-56623637	*Fbxo8/Hand2*
chr8:71687314-71688316	*Insl3/Jak3*
chr8:80917179-80918973	*Gab1/Usp38**
chr8:94153505-94154175	*Mt3**
chr8:105374715-105374834	*Plekhg4/Slc9a5*
chr8:109075175-109075765	D030068K23Rik
chr8:117109911-117110338	*Bco1*
chr8:121273267-121273661	*Foxl1*/1700018B08Rik
chr8:121541793-121542193	1700018B08Rik
chr8:121542486-121543053	170018B08Rik/30M09Rik
chr8:122864817-122865344	*Cdh15*
chr8:124363052-124363447	*Pgbd5/Galnt2**
chr8:126476627-126476857	*Gm17296/Irf2bp2*
chr9:3199699-3199906	4930433N12Rik
chr9:20857639-20859264	A230050P20Rik/*Rdh8**
chr9:20911944-20912365	*Dnmt1*
chr9:45115913-45116264	*Scn2b*/Gm10684
chr9:89737795-89738762	*Ankrd34c**
chr9:89879711-89881749	***Rasgrf1*/4930524008Rik***
chr9:91380349-91381121	*Zic4*
chr10:7614884-7617224	*Lrp11*
chr10:13090448-13091600	***Plagl1****
chr10:79735189-79735518	*Polrmt/Hcn2/Bc1**
chr11:4440809-4440945	*Hormad2*
chr11:5516696-5516994	*Xbp1/Znrf3**
chr11:12025737-12025932	***Grb10****

***** Indicates that NORED region overlaps bimodal (MethylMosaic) region
Bold indicates NORED region overlaps known imprinted germline DMR

Table 2.1 (continued) Identification, demarcation, and characterization of highest ranked NORED

Chromosome location	Nearest gene(s)
chr11:12025971-12026410	***Grb10****
chr11:12026426-12026880	***Grb10****
chr11:22007227-22007753	*Otx1/Ehbp1**
chr11:22971545-22974429	***Zrsr1/Commd1****
chr11:54807023-54807079	*Lyrm7os/Cdc42se2*
chr11:58961910-58961988	*Trim17*
chr11:67084007-67084273	*Myh3*
chr11:115441364-115441906	*Trim80**
chr11:117479477-117479864	Gm11733/*Sept9*
chr11:117790191-117790485	6030468B19Rik/*Tmc8*
chr11:119258377-119259019	*Gaa/Ccdc40*
chr11:120316425-120316754	*Actg1/Bahcc1*
chr11:120949785-120950433	*Slc16a3*
chr11:121519413-121520621	*Zfp750/Tbcd**
chr12:71577311-71577941	*4930404H11Rik/Daam1**
chr12:84640456-84641612	*Vrtn*
chr12:104448190-104448500	*Gsc/Serpina3n*
chr12:109524554-109526114	***Meg3/Dlk1****
chr12:109527558-109529443	***Meg3/Dlk1****
chr12:109529532-109531222	***Meg3/Dlk1****
chr12:109539200-109539966	*Meg3/Dlk1**
chr12:109541027-109541475	*Meg3**
chr12:109541501-109542870	*Meg3**
chr13:12833830-12837749	*Prl2c3/Prl2c2*
chr13:13450125-13451423	*Nid1*
chr13:23285704-23286065	4933404K08Rik
chr13:23286248-23286415	4933404K08Rik
chr13:23299932-23300187	4933404K08Rik
chr13:23301151-23301633	4933404K08Rik
chr13:23306774-23307322	4933404K08Rik
chr13:23313146-23313412	4930557F10Rik
chr13:30947041-30947819	*Hus1b/Exoc2**
chr13:47013723-47014359	*Nhlrc1*
chr13:52928754-52929435	*Auh*
chr13:53194451-53194648	*Ror2*
chr13:56522315-56522451	*Fbxl21/ll9*
chr13:84236742-84237442	*Tmem161b*
chr13:104531878-104532411	*Adamts6/Cwc27*
chr13:120024970-120025495	B020031M17Rik/Gm20767
chr13:120026729-120027576	B020031M17Rik/Gm20767
chr13:120028688-120029477	HB020031M17Rik/Gm20767*
chr13:120030479-120030889	GM20767/B020031M17Rik

***** Indicates that NORED region overlaps bimodal (MethylMosaic) region
Bold indicates NORED region overlaps known imprinted germline DMR

Table 2.1 (continued) Identification, demarcation, and characterization of highest ranked NORED

Chromosome location	Nearest gene(s)
chr13:120033528-120033703	Gm21188/B020031M17Rik
chr13:120035672-120036958	Gm21188*
chr14:58073331-58073940	*Fg19*
chr14:65404585-65404937	*Pnoc*
chr14:68122518-68124228	A230070E04Rik/*Nefm*
chr14:68124336-68124471	*Nefm*/ A230070E04Rik
chr14:79772279-79772640	*Pcdh8*/Gm10845
chr15:27476183-27476613	*Ank*
chr15:72809195-72811003	***Peg13/Trappc9****
chr15:78037646-78038392	*Cacng2*
chr15:79972322-79972965	*Cbx7/Pdgfb*
chr15:85131470-85131874	*Smc1b*
chr16:11035880-11036418	*Snn*/Gm4262
chr16:20732118-20732136	*Chrd/Thpo*
chr16:22405637-22406912	*Etv5**
chr16:22407226-22408474	*Etv5**
chr16:22410190-22411163	*Etv5**
chr17:5958515-5959056	*Synj2*
chr17:6582455-6582783	*Dynlt1c*
chr17:12741566-12742455	***Airn/Igf2r****
chr17:12742478-12742859	***Airn/Igf2r****
chr17:21383671-21383999	*Zfp677*
chr17:26603052-26603573	*Ergic1**
chr17:28858375-28858617	*Pnpla1*
chr17:30874696-30874978	*Glp1r/Dnah8*
chr17:57492974-57493278	*Vmn2r120/Emr1*
chr17:83874898-83875201	*Haao*/4933433H22Rik
chr18:12972131-12975659	***Impact****
chr18:24671762-24671989	*Mocos*
chr18:25478389-25478493	*Celf4*
chr18:36940664-36940927	*Pcdha2/Pcdha1*
chr18:36992516-36992804	*Pcdha8/Pcdha1*/Gm37013
chr18:37012737-37013324	*Pcdha11/Pcdha1*/Gm37013
chr18:37402798-37403276	*Pcdhb9*/Gm37013/Gm38666
chr18:65698431-65698803	*Oacyl*
chr19:16035301-16036310	*C130060C02Rik/Gnaq*
chr19:61225200-61225451	*Csf2ra*

***** Indicates that NORED region overlaps bimodal (MethylMosaic) region

Bold indicates NORED region overlaps known imprinted germline DMR

MethylMosaic: genome-wide detection of bimodal methylation

As described above, identified NORED regions include known imprinted gASMs, which have bimodal methylation patterns. To characterize genomic regions that have potential to exhibit allelic methylation patterns, we sought a genotype-independent approach that could be applied in homozygous WT ESCs such as J1 ESC line. The bimodal distribution of methylation patterns has long been used in conventional bisulfite Sanger sequencing to validate ASMs. Stimulated by the concept and experimental design, we implemented a computational program, MethylMosaic, to exploit the bimodal methylation patterns that are characteristic of imprinted ASMs. Notably, approaches based on similar concept have successfully identified ASMs (Fang et al., 2012; Peng and Ecker, 2012).

To identify bimodal regions by 'MethylMosaic,' we first calculated the read-level methylation and then the proportion of hypomethylated reads (hypomethylation index) and the proportion of hypermethylated reads (hypermethylation index) around each CpG site within the mouse genome. We then determined the true positive rate (TPR; 21 well-characterized gASMs considered as true positives) for various cutoffs of hypo- and hypermethylation indices to identify bimodal CpG sites (see details in Methods). To calculate the FPR, we simulated ten null datasets by shuffling the methylation calls among reads at CpG positions. Importantly, the randomization of methylation among reads at each CpG site has the potential to alter read-level methylation, but keeps the CpG-level methylation intact. We calculated hypo- and hypermethylation indices and applied cutoffs (as described above) to identify bimodal CpG sites for null datasets, which were considered false positives. The FPR for each cutoff range was determined by averaging the FPR from null datasets. Based on the ROC curve, we selected the range from 0.2 to 0.75 as the bounds for hypermethylation and hypomethylation indices (Supplemental Fig. S2.3A). For the WT dataset and the null datasets, we clustered individual CpG sites into regions and ranked regions by the number of CpG sites. Null datasets were used to determine region-level FDR.

MethylMosaic analyses identified 2487 regions as bimodal at FDR = 0.20 (Fig. 2.3A, Supplemental Fig. S2.3B, and Supplemental Table S2.3). Consistent with NORED, neither *Slc38a4* nor *Gnas 1A* were identified as bimodal, presumably due to low methylation in WT (Fig. 2.1A). All remaining 19 established gASMs were among the highest 32 ranked regions (FDR < 3.85×10^{-3}; Fig. 2.3A–B and

Supplemental Fig. S2.3B). Of *Peg3/Usp29* ASM, 166 consecutive CpG sites met criteria for hyper- and hypomethylation index cutoffs. From overlapping reads of CpG-centered windows, we retrieved 674 reads covering 169 CpG sites; 427 of those reads contained at least three CpG sites per read (Fig. 2.3C). Compared to 20–40 PCR clones in Bisulfite Sanger sequencing, hundreds of reads (up to 674 reads here) demonstrate the robustness of MethylMosaic. MethylMosaic also revealed bimodal distribution of hyper- and hypomethylated at *Gipr*/*Eml2*, *Hcn2*/*Polrmt*, *Errfi1*/*Park7*, and *Hus1b*/*Exoc2* (Fig. 2.3D–G). At these four loci, we retrieved 151, 126, 168, and 83 reads having at least three CpG sites from 248, 163, 254, and 146 total reads, respectively.

Characterization of genomic loci that are NORED and have bimodal methylation

As expected, both NORED and MethylMosaic identified 19 well-characterized gASMs. To characterize the extent to which other NORED regions have potential to exhibit allelic methylation patterns, we compared approximately 2500 regions presented for each method. Comparison demonstrated that only 8.4 percent (207) of 2468 NORED regions overlapped at least one bimodal region and only 2.2 percent (152) of 2487 bimodal regions overlapped at least one NORED region (Supplemental Fig. S2.4A–B). Therefore, the majority of NORED regions were not bimodal (Supplemental Fig. S2.4A).

To rule out the possibility that the low rate of NORED that are also bimodal is driven by the number of regions presented for each method, we determined the proportion of bimodal NORED regions at multiple FDRs for MethylMosaic. For 2468 NORED presented in Supplemental Table S2.2, bimodal NORED regions would not become the majority of total NORED regions until MethylMosaic FDR = 0.74. For the 207 NORED (FDR = 0.005) regions presented in Table 2.1, 50.2 to 71.5 percent would be considered bimodal between FDR = 0.42 and FDR = 0.90 for MethylMosaic. We therefore conclude that NORED and MethylMosaic are complementary, but independent methods for identifying genomic regions with different characteristics. That is, gASM are both NORED and bimodal; however, there are indeed other genomic regions that are either NORED or bimodal, but not both.

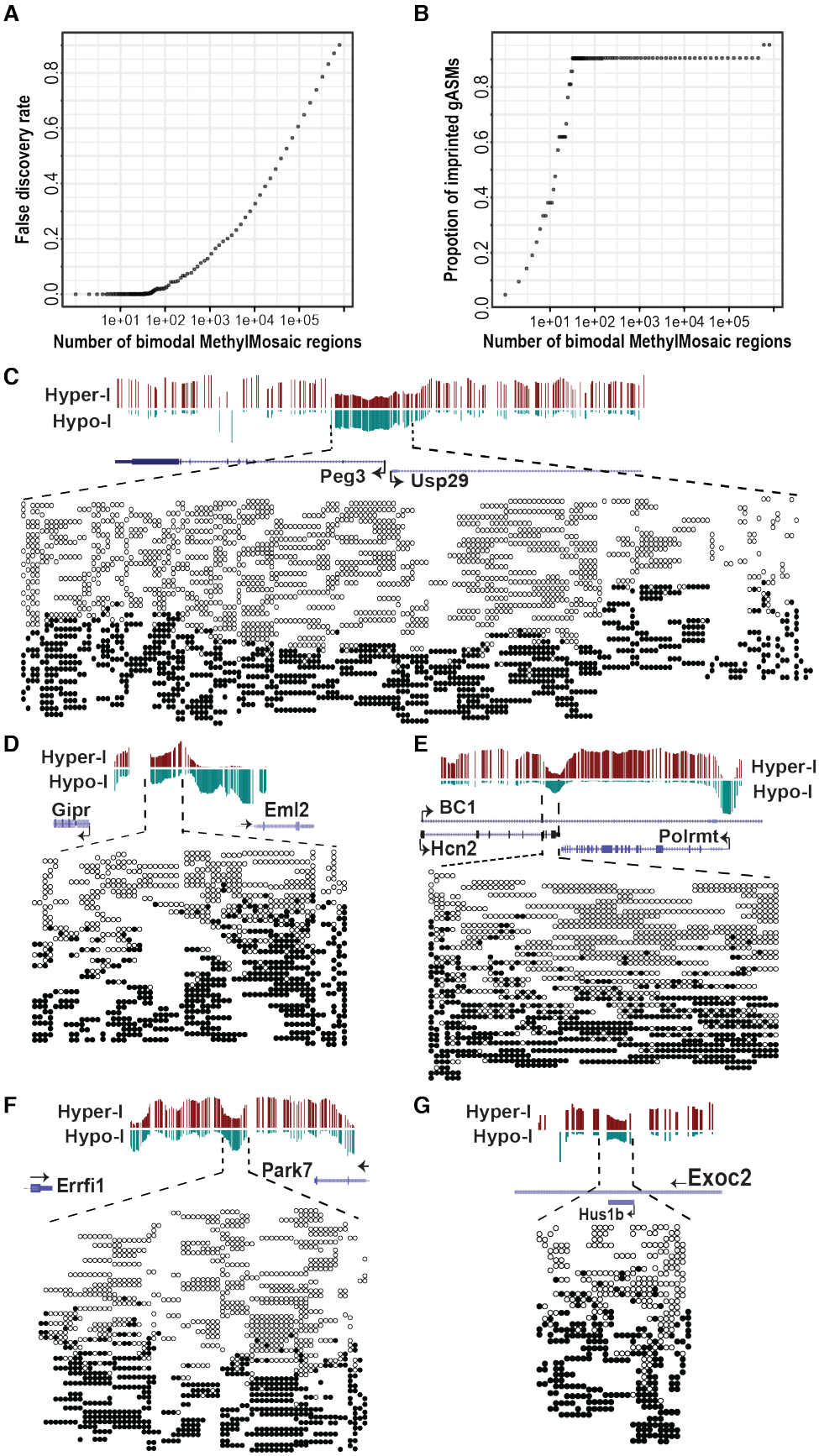
Of the top 207 ranked NORED regions presented in Table 2.1, 75 (36.2%) were bimodal. In addition to gASMs, NORED regions that had corresponding bimodal regions included *Gipr*/*Eml2*,

Hcn2/*Polrmt*, *Errfi1*/*Park7*, and *Hus1b*/*Exoc2* (Fig. 2.3D–G). Note, *Hus1b* ASM is within the intron of *Exoc2* gene. NORED regions at possible gASM of *Nnat/Blcap* and imprinted *Gab1* were also identified as bimodal, whereas potential gASM at *Cdh15* was not (Table 2.1).

Characterization of genes associated with regions that are both NORED and bimodal, exclusively NORED, or exclusively bimodal

Using Chemical and Genetic Perturbations (CGP) gene sets from MSigDB (see Methods), we first asked whether genes associated with 2468 NORED regions and genes associated with 2487 bimodal regions were enriched in imprinted genes. NORED regions were enriched for imprinted genes (27 genes, $q = 2.03 \times 10^{-21}$), whereas bimodal genes did not report imprinted genes within the top 100 enriched gene sets (Supplemental Tables S2.4–S2.5). MethylMosaic regions were enriched for the gene set of high CpG density promoters bearing both H3K4me3 and H3K27me3 histone modifications (167 genes, $q = 1.80 \times 10^{-53}$; Supplemental Table S2.5). Notably, for genes associated with 207 regions considered both NORED and bimodal, imprinted genes was the top gene set identified (19 genes, $q = 5.84 \times 10^{-31}$; Supplemental Table S2.6).

Figure 2.3 Read-level methylation reveals the bimodal distribution of hyper- and hypomethylated reads at gASMs and other NORED loci. (A) False discovery rate of MethylMosaic. (**B)** 19 gASMs rank among top 32 MethylMosaic regions. **(C)** MethylMosaic of known *Peg3* gASM demonstrates the robustness of the approach. *Peg3* has 674 reads, of which 427 have at least three CpG sites. **(D–G)** MethylMosaic analyses reveals bimodal distribution of hyper- and hypomethylated reads at NORED regions that are not established gASMs. **(D)** *Gipr*/*Eml2* has 248 reads, of which 151 have at least three CpG sites. **(E)** *Hcn2/Polrmt* has 163 reads, of which 126 have at least three CpG sites. **(F),** *Errfi1*/*Park7* and has 254 reads, of which 168 have at least three CpG sites. **(G)** *Hus1b*/*Exoc2* has 146 reads, of which 83 have at least three CpG sites. **(C–G)** Hypermethylation Index is abbreviated as Hyper-I and Hypomethylation Index is abbreviated as Hypo-I (see Methods).


A
False discovery rate
0.0
0.2
0.4
0.6
0.8
1e+01 1e+02 1e+03 1e+04 1e+05
Number of bimodal MethylMosaic regions
B
Propotion of imprinted gASMs
0.0
0.2
0.4
0.6
0.8
1e+01 1e+02 1e+03 1e+04 1e+05
Number of bimodal MethylMosaic regions
C
Hyper-l
Hypo-l
Peg3
Usp29
D
Hyper-l
Hypo-l
Gipr
Eml2
E
Hyper-l
Hypo-l
BC1
Polrmt
Hcn2
F
Hyper-l
Hypo-l
Errfi1
Park7
G
Hyper-l
Hypo-l
Exoc2
Hus1b

To provide insight into potential functional implications for regions detected separately or by both NORED and MethylMosaic, we identified enrichment of CGP gene sets for 123, 1427, and 1627 gene identifiers associated with 207 bimodal NORED regions, 2261 exclusively NORED regions, and 2335 exclusively MethylMosaic regions, respectively. Other than imprinted genes, disease-relevant gene sets identified for bimodal NORED-associated genes included nasopharyngeal carcinoma (22 genes, $q = 5.28 \times 10^{-6}$), genes upregulated in mutated KRAS lung cancer model (10 genes, $q = 7.69 \times 10^{-4}$), pancreatic cancer (6 genes, $q = 4.05 \times 10^{-3}$), TP53 targets (11 genes, $q = 2.24 \times 10^{-2}$), Alzheimer's disease upregulated genes (13 genes, $q = 3.07 \times 10^{-2}$), female fertility (3 genes, $q = 9.22 \times 10^{-3}$) and metabolic syndrome (11 genes, $q = 2.53 \times 10^{-2}$; Supplemental Table S2.6).

NORED exclusive-associated genes were also enriched nasopharyngeal carcinoma (163 genes, $q = 1.92 \times 10^{-31}$), TP53 targets (113 genes, $q = 6.18 \times 10^{-24}$), Alzheimer's disease upregulated genes (146 genes; 1.98×10^{-26}), and metabolic syndrome (99 genes, $q = 4.06 \times 10^{-16}$; Supplemental Table S2.7). Enrichment for genes hypermethylated in liver cancer (96 genes, $q = 4.96 \times 10^{-22}$), lung cancer (43 genes, $q = 5.41 \times 10^{-11}$), and lymphoma tumors of transgenic mice (16 genes, $q = 9.23 \times 10^{-10}$) was observed only in NORED exclusive-associated genes (Supplemental Table S2.7). Furthermore, NORED exclusive-associated genes were uniquely enriched in genes characterized by H3K27me3 with polycomb proteins (SUZ12 or EED) bound to promoters that experience *de novo* DNA methylation in cancers (18 genes, $q = 6.94 \times 10^{-9}$; Supplemental Table S2.7).

Similar to bimodal NORED and NORED exclusive regions, MethylMosaic exclusive-associated genes were also enriched nasopharyngeal carcinoma (149 genes, $q = 1.07 \times 10^{-19}$), TP53 targets (126 genes, $q = 3.13 \times 10^{-26}$), Alzheimer's disease upregulated genes (205 genes, $q = 9.61 \times 10^{-52}$), and metabolic syndrome (125 genes, $q = 1.31 \times 10^{-24}$; Supplemental Table S2.8). Enrichment for genes upregulated in chronic myleogenous leukemia (151 genes, $q = 2.07 \times 10^{-32}$) and upregulated in uveal melanoma (104 genes, $q = 8.85 \times 10^{-29}$) were uniquely identified for MethylMosaic exclusive regions (Supplemental Table S2.8).

Independent characterization of genomic regions with bimodal methylation patterns with newly generated ESC lines

We next aim to exploit regions with bimodal methylation patterns with experiments. Four scenarios could explain a genomic region bearing bimodal methylation patterns: (1) bona fide imprinted ASM (i.e., parent-of-origin dependent and genetics-independent); (2) genetics-dependent ASM but independent of parent-of-origin; (3) one allele with hypomethylation (or hypermethylation) in half of cells and the same allele with hypermethylation (or hypomethylation) in the remaining half of cells (named as switchable ASM); and (4) half of cells with biallelic hypermethylation and half of cells with biallelic hypomethylation (see illustration below).

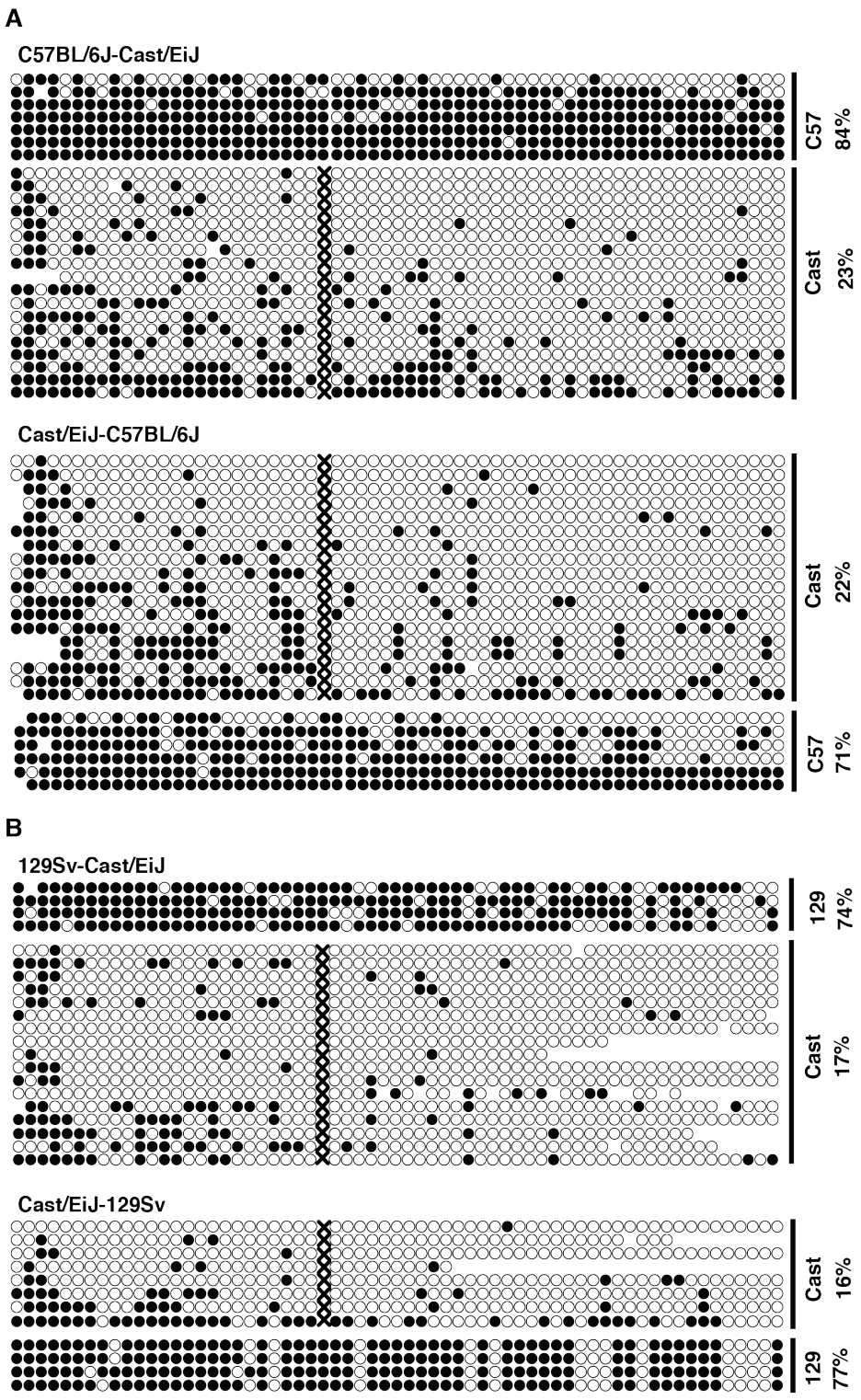
To experimentally exploit these identified MethylMosaic regions in the mouse genome, we sought hybrid ESC lines with single nucleotide polymorphisms (SNPs) between two alleles for characterizations. We are interested in the characterization of genes with potential in neurological disorders. Therefore, we focused our validation on two ASMs: *Hcn2* with known roles in epilepsy, inflammatory and chronic pain (Emery et al., 2011), and *Park7* (or *DJ-1*) for Parkinson's disease (Bonifati et al., 2003). We first used several available ESC lines; however, we did not detect the bimodal methylation patterns (data not shown). We reason that because these lines were generated many years ago, multiple passages might result in aberrant methylation patterns similar to loss of imprinting described in human ESCs (Rugg-Gunn et al., 2007). In support of our reasoning, *H19*/*Igf2* ASM was frequently lost in ESC lines (confirmed, data not shown). Alternatively, ASMs might be transient (Babak et al., 2015), and the developmental stage of inner cell mass might not be appropriate for ASMs. We therefore decided to generate our own mouse ESC lines from F1 hybrid mice (129S1/SvimJ and Cast/EiJ or C57BL/6J and Cast/EiJ).

With DNA from two ESC lines (C57Cast and CastC57), we did bisulfite Sanger sequencing to examine methylation status of *Hcn2/Polrmt* ASM. We tried and succeeded with one pair of primers that cover 62 CpG sites, thereby enabling us to have a better insight of *Hcn2/Polrmt* ASM (Fig. 2.4A). Indeed, our data revealed allele-specific DNA methylation patterns: Cast allele (paternal) from C57Cast line was hypomethylated (23% methylation), whereas C57 allele (maternal) was hypermethylated (84% methylation) (Fig. 2.4A, top panel). Consistently, Cast allele (maternal) from CastC57 line was

hypomethylated (22% methylation), whereas C57 allele (paternal) was hypermethylated (71% methylation) (Fig. 2.4A, bottom panel). While bimodal methylation patterns are expected according to our MethylMosaic analyses (Fig. 2.3E), it is unexpected that this allele-specific methylation pattern is independent of parent-of-origin. Whether Cast allele was paternal or maternal copy, it was always hypomethylated.

To reinforce our confidence with our results above, we generated two more ESC lines from strains 129 and Cast, as biological replicates. Indeed, bisulfite Sanger sequencing data confirmed the allelic methylation patterns. Within 129Cast ESC line, 129 allele (maternal) was hypermethylated (74%), and Cast allele (paternal) was hypomethylated (17%); Within Cast129 ESC line, Cast allele hypomethylated (16%) and 129 hypermethylated (77%) (Fig. 2.4B). Again, the allele specificity of *Hcn2* ASM was independent of parent-of-origin. Therefore, we conclude that *Hcn2* ASM has indeed bimodal methylation patterns, as revealed by MethylMosaic data in J1 ESC line and bisulfite Sanger sequencing data in four independent ESC lines (129Cast, Cast129, C57Cast, and CastC57), and that Cast allele always has the hypomethylated CpG sites, independent of parent-of-origin. Because of the latter, *Hcn2* ASM is not a bona fide imprinted ASM.

Figure 2.4 Bisulfite Sanger sequencing with four F1 hybrid ESCs confirmed the AMS at Hcn2/Polrmt locus. (A–B) The "X" indicates a CpG site not present in allele. **(A)** The paternal Cast allele of ESC line from cross between C57 and Cast (top) was hypomethylated; the maternal Cast allele of ESC line from cross between Cast and C57 was hypomethylated (bottom). **(B)** The paternal Cast allele of ESC line from cross between 129 and Cast (top) was hypomethylated; the maternal Cast allele of ESC line from cross between Cast and 129 was hypomethylated (bottom). **(C)** DNA sequence alignment revealed a SNP-introduced binding motif of transcription factor MTF-1.


A
C57BL/6J-Cast/EiJ
C57 84%
Cast 23%
Cast/EiJ-C57BL/6J
Cast 22%
C57 71%
B
129Sv-Cast/EiJ
129 74%
Cast 17%
Cast/EiJ-129Sv
Cast 16%
129 77%
C
129S1/SvlmJ GCCTCTCTTCCAACTTGTGACCCTTGaGCGCCGCCCCGCGGGCCGGGCGGGGCCGTCATC
C57BL/6NJ GCCTCTCTTCCAACTTGTGACCCTTGaGCGCCGCCCCGCGGGCCGGGCGGGGCCGTCATC
CAST/EiJ GCCTCTCTTCCAACTTGTGACCCTTGcGCGCCGCCCCGCGGGCCGGGCGGGGCCGTCATC
MTF-1

One SNP variant of the Cast allele results in a binding motif for transcription factor at Hcn2/Polrmt ASM

The hypomethylated Cast allele promoted us to ask whether this biased hypomethylation was related to genetic variations. There are three SNP variants around *Hcn2* ASM. One variant (cytosine in rs240718423 in Cast allele), intriguingly, results in an additional CpG dinucleotide. The resulting sequence of TG***C***GCGC becomes the core consensus sequence TGCRCNC (R=A or G, N=any nucleotide) of a metal regulatory transcription factor MTF-1 (Fig. 2.4C). MTF-1 is a pluripotent regulator that regulates cell adaptation to various stress conditions (primarily exposure to heavy metal, and stresses of hypoxia and oxidative stress) (Selvaraj et al., 2005; Sims et al., 2012). In contrast, TG***A***GCGC in the allele of 129 or C57 is not a binding motif for MTF-1. Whether this MTF-1 predisposes Cast allele to low methylation during cycles of demethylation and methylation for the generation of Hcn2 ASM, however, remains to be determined. Another variant (adenine in rs259784301 in Cast allele), results in one less CpG dinucleotide in Cast than in 129 or C57 alleles (Fig. 2.4A–B).

The difference of multiple SNP variants may predispose the Cast allele to hypomethylation, prompting us to examine the similar possibility of corresponding regions in J1 (inbred 129) ESCs. We did Sanger sequencing of PCR amplified products. As expected, the corresponding region did not contain any variants at known SNP positions or *de novo* mutations (Supplemental Fig. S2.5), ruling out the possibility of genetic variants in J1 ESCs at this locus. Therefore, we conclude that genetic variations are not necessary to render this *Hcn2* region with bimodal methylation patterns, but the SNP-associated motif introduced in Cast allele may expose it to hypomethylation (Fig. 2.4).

Independent validations of Park7 ASM reveal a random, switchable ASM

Simultaneously, we have examined the *Park7* ASM (see bimodal methylation in Fig. 2.3F) in four new ESC lines. The primers we used cover 18 CpG sites, and methylation status at paternal allele or maternal allele showed interesting patterns: Out of 35 PCR clones examined for CastC57 ESC line, 22 clones were for methylation level (49%) of maternal Cast allele; 13 clones, for methylation level (54%) of paternal C57 allele (Fig. 2.5B). Intriguingly, half of clones from either Cast allele or C57 allele were hypermethylated or hypomethylated. Out of 24 PCR clones from an independent C57Cast ESC line, 12

clones were for methylation level (59%) of maternal C57 allele; the remaining 12 clones, for methylation level (43%) of paternal Cast allele. Consistently, about half of 12 clones were either hypomethylated or hypermethylated (Fig. 2.5C). Bisulfite Sanger sequencing data from two biological replicates (two ESC lines; Cast129 and 129Cast) reproduced observations above (data not shown). These data suggested a random, switchable allele-specific methylation pattern.

Compared to *Hcn2* ASM, the genetic differences between Cast allele and C57 or 129 allele of *Park7* ASM did not result in a new binding motif for a potential transcription factor (data not shown), thereby not predisposing one allele for hypomethylation. Without a new binding motif, genetic variations of *Park7* ASM presumably behave similarly with the *Hcn2* ASM in J1 ESC line. The latter has no genetic variants. Collectively, we conclude that *Park7* ASM indeed shows bimodal methylation pattern, and that the differences between Cast allele and C57 (or 129) allele did not result in a preference of one allele for hypomethylation.

Conserved ASM at Hcn2/Polrmt locus in the human genome

Having demonstrated bimodal methylation patterns of *Hcn2/Polrmt* ASM in five independent mouse ESC lines (J1, 129Cast, Cast129, CastC57, and C57Cast), we next explored the evolutionarily conserved ASM at *Hcn2*/*Polrmt* locus in the human genome for further validation. Because bimodal patterns were detected in five mouse ESC lines, but a recent examination of cortical neurons from the cross (129 x Cast) did not find ASM (Xie et al., 2012), we decided to examine human ESC line. In the human genome, the human *Hcn2* and *Polrmt* genes have convergent genomic organization, similar to the mouse orthologs (Fig. 2.6). We designed two pairs of primers with one inside of the predicted ASM and the other outside (Fig. 2.6). Strikingly, monoclonal sequencing reads with 45 CpG sites from the 'inside' pairs have roughly half hypo- and half hypermethylated reads (Fig. 2.6A). In contrast, all PCR clones from the 'outside' pairs contain hypermethylated reads (Fig. 2.6B). We conclude that *Hcn2* ASM is conserved in the human genome.

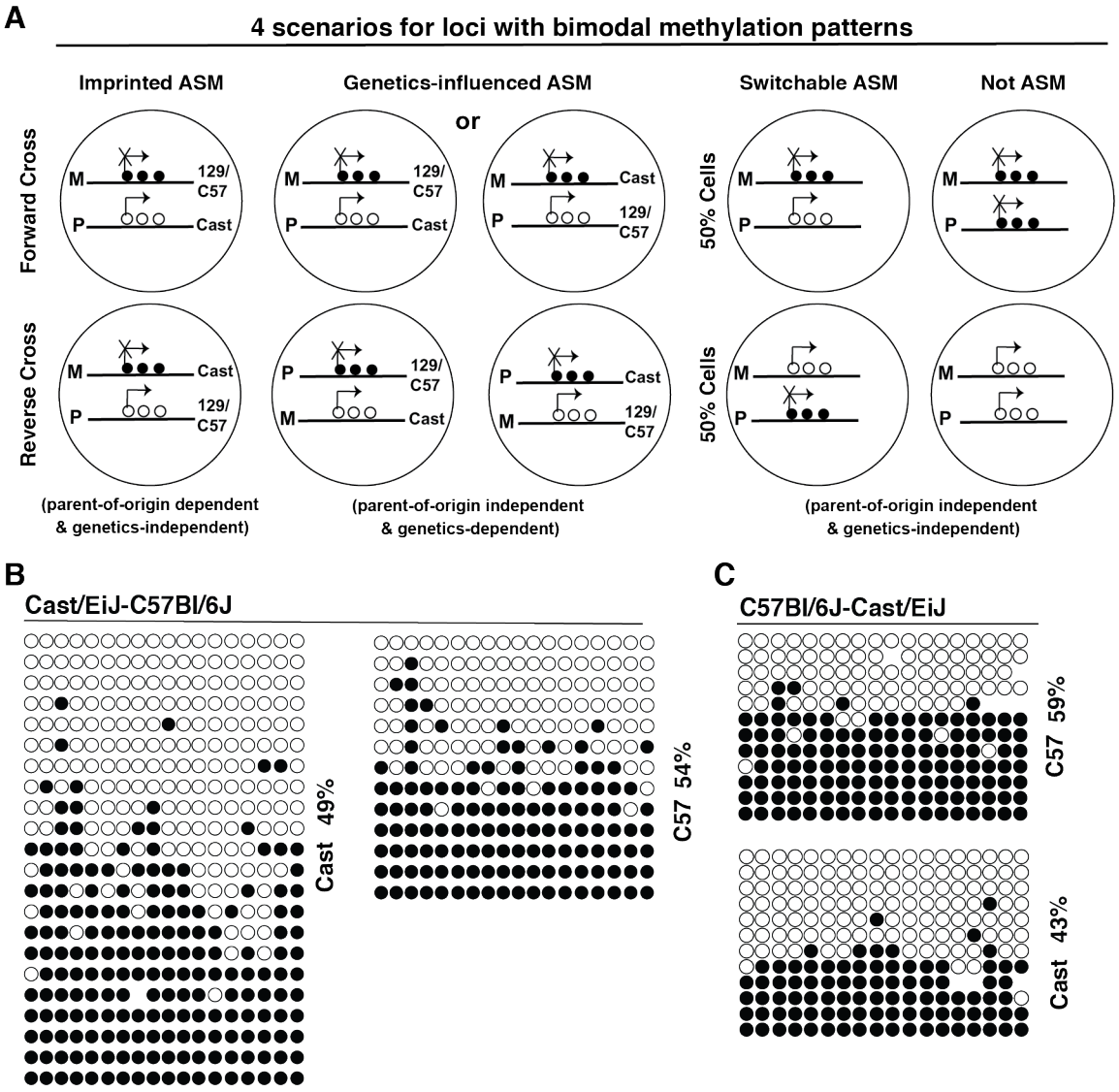


Figure 2.5 Four scenarios for regions with bimodal methylation patterns and Park7 ASM. (A) Four scenarios include: bona fide imprinted ASM; genetics-dependent ASM but independent of parent-of-origin; one allele with hypomethylation (or hypermethylation) in half of cells and the same allele with hypermethylation (or hypomethylation) in the remaining half of cells; and half of cells with biallelic hypermethylation and half of cells with biallelic hypomethylation. **(B)** Half of PCR clones were either hypermethylated or hypomethylated from the maternal Cast allele from CastC57 ESC line (left); similarly, half of PCR clones were either hypermethylated or hypomethylated from the paternal allele from CastC57 ESC line. **(C)** Half of PCR clones were either hypermethylated or hypomethylated from the maternal C57 allele from C57Cast ESC line (left); similarly, half of PCR clones were either hypermethylated or hypomethylated from the paternal Cast allele from C57Cast ESC line.

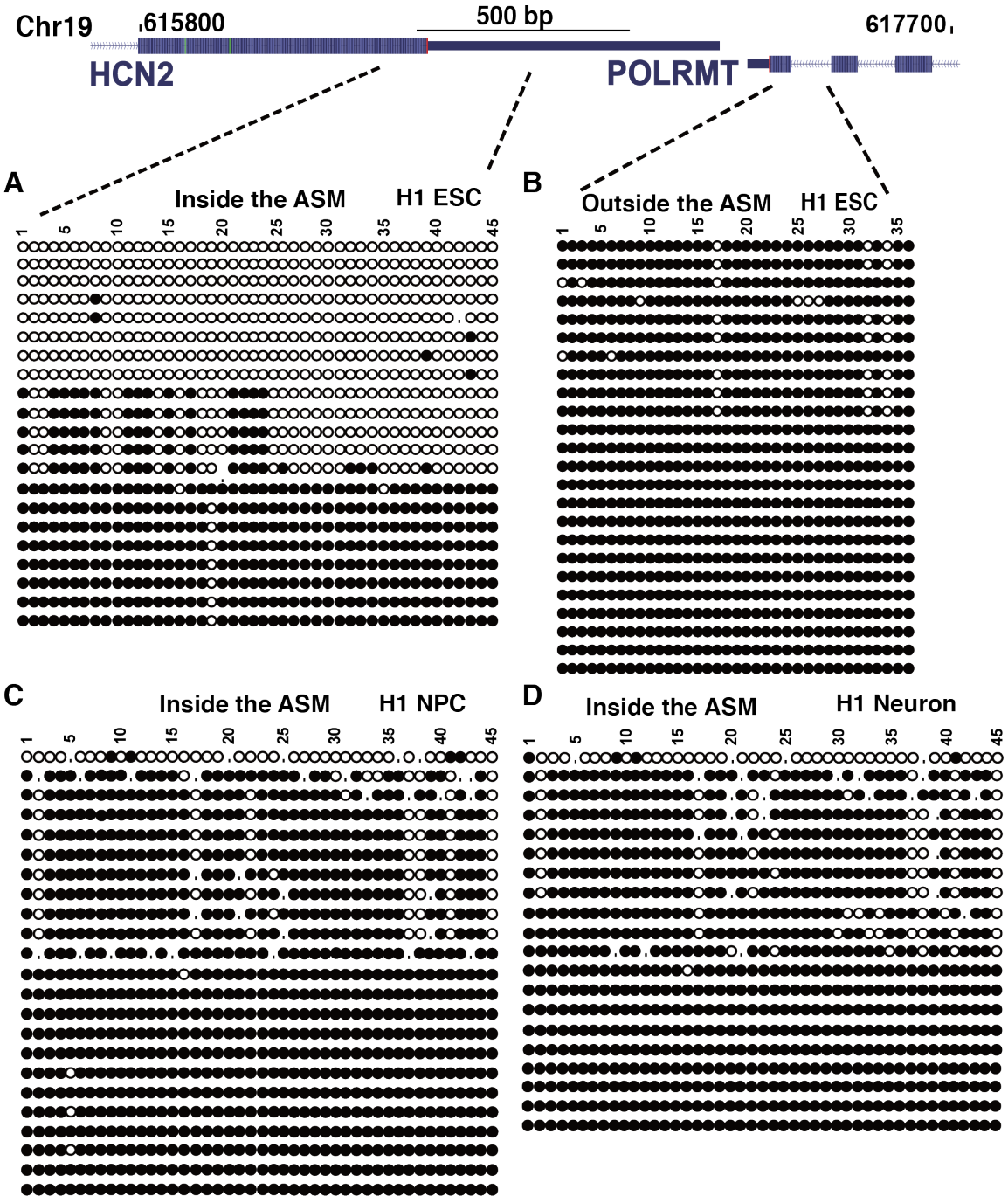


Figure 2.6 Bisulfite Sanger sequencing confirmed that transient DMR at *Hcn2/Polrmt* locus is conserved in the human genome. (A) Bisulfite Sanger sequencing revealed bimodal distribution of hyper- and hypomethylated reads. **(B)** Bisulfite Sanger sequencing revealed fully methylated reads from a neighboring control region. **(C–D)** Bisulfite Sanger sequencing revealed a *de novo* methylation process to cause hypermethylation of CpG sites from the hypomethylated allele upon differentiation of H1 ESCs into NPCs **(C)** or neurons **(D)**.

The transient ASM at Hcn2/Polrmt locus participates in early embryonic development

The observation that cortical neurons did not have a bimodal methylation pattern at *Hcn2*/*Polrmt* locus (Xie et al., 2012) (data not shown), suggests a role of *Hcn2* ASM during development. We then examined *Hcn2* ASM using *in vitro* differentiated neuron progenitor cells (NPCs) and neurons derived from H1 ESCs. Consistent with mouse cortical neurons, human NPCs and neurons became fully methylated, confirming the loss of the bimodal patterns at *Hcn2*/*Polrmt* locus upon differentiation (Fig. 2.6C–D). Therefore, this ASM is transiently presented during early embryonic development.

3. Discussion

DNMT1-depenent methylation regions for mechanistic insights of disease susceptibility

As the predominantly expressed DNMT especially in somatic cells, DNMT1 is the favorite enzyme out of three DNMTs for investigation (for example, the development of DNMT1 inhibitor in the treatment of cancers), and DNMT1-dependent methylation patterns are presumably important for understanding disease pathogenesis. Herein we have used/developed two approaches to characterize DNMT1-dependent methylation patterns at genomic regions in the mouse genome. The NORED approach, which is a newly developed method in this study, identified 2468 genomic regions dependent on DNMT1 function. Among them, 207 regions also show bimodal methylation patterns (i.e., also MethylMosaic regions). Regions showing bimodal methylation patterns include 19 known imprinted gASMs. Relevant to human health, these gASMs and tissue-specific ASMs are particularly vulnerable to environment-induced perturbation (Jirtle and Skinner, 2007; Murphy and Hoyo, 2013; Susiarjo et al., 2013). Given the facts that gASMs and novel MethylMosaic regions share the feature of allelic methylation patterns, it is reasonable to expect that these MethylMosaic regions are also vulnerable to environmental exposure. Indeed, unpublished base resolution methylomes of endocrine disruptor-exposed mice reveal many such regions were impacted. We therefore expect that these NORED/MethylMosaic regions will be used extensively to inform studies in exposed mice in the future.

Many genes from both NORED and MethylMosaic regions have potential in human diseases. *Hcn2* (Fig. 2.3E, 2.4, 2.6) is responsible for hyperpolarization-activated cation (HCN) channel, which is

linked to the generation of cardiac pacemaker depolarization and the control of neuronal excitability and plasticity. *Hcn2* is implicated in the pathogenesis of epilepsy (Chung et al., 2009; Ludwig et al., 2003) and linked to chronic pain as well (Emery et al., 2011). *Polrmt* is for mitochondrial transcription (Wanrooij et al., 2008). Additional neurological disorders, including Parkinson's disease at *Park7*(DJ-1)/*Errfi1* locus (Fig. 2.3F, 2.5) (Bonifati et al., 2003; Dawson et al., 2010). In addition to neurological activities, novel DMRs have potential roles in type 2 diabetes at the *Gipr*/*Eml2* locus (Fig. 2.2E, 2.3D) (Saxena et al., 2010) and in cell cycle checkpoint at *Hus1b*/*Exoc2* locus (Fig. 2.3G) (Hang et al., 2002) were identified. Altogether, our data open new windows to markedly improve the understanding of many complex human diseases (Piedrahita, 2011; Skaar et al., 2012; Smith et al., 2006; Weksberg, 2010).

While bimodal regions were not reported as being enriched for imprinted genes (not within the top 100 enriched gene sets from CGP), they were enriched for genes identified as having both active H3K4me3 and inactive H3K27me3 histone modifications (within the top 5 enriched gene sets; Supplemental Table S2.5). This would be consistent with a scenario of allelic methylation, with an unmethylated allele having active histone marks and a methylated allele having inactive histone marks. Alternatively, it could be explained by a mixed cell population, with half of cells having active epigenetic marks and the other half having inactive marks. Notably, DNMT1-dependent regions were enriched for imprinted genes (within the top 20 enriched gene sets; Supplemental Table S2.4). Overall our results indicate that MethylMosaic identified bimodal regions, but imprinted genes are more obvious in NORED.

Further reinforcing the importance of recognizing DNMT1-dependent regions for disease, NORED exclusive regions uniquely identified genes that are known to become *de novo* DNA methylated in cancer and genes that have previously been reported as hypermethylated in cancer (Supplemental Table S2.7). Perhaps reflecting the complexity of neurodegenerative and metabolic disorders, NORED-exclusive, MethylMosaic-exclusive, and bimodal NORED were all enriched for Alzheimer's upregulated genes and metabolic syndrome network genes (Supplemental Tables S2.6–S2.8). Importantly, bimodal NORED regions were highly enriched in imprinted genes (highest ranked gene set) and uniquely enriched in female fertility genes, both of which could have implications 'developmental origin of adult disease' and 'transgenerational epigenetic inheritance' (Supplemental Table S2.6). Enrichment analysis of CGP

reaffirmed the assertion that allelic methylation and DNMT1-depenence are separate characteristics that coincide at gASMs and additional bimodal NORED regions.

NORED demarcates DNMT1-dependent methylation at ASMs

Knowing the exact genomic region of an ASM is important to diagnose and understand the pathogenesis of imprinting disorders, such as Prader-Willi and Angelman syndromes (Kim et al., 2012; Skryabin et al., 2007). Prior examination of patient samples with microdeletions to define the location of ASMs is time consuming and limited by the availability of patient samples (Ohta et al., 1999). Systematic identification of all ASM sizes and characterization of alterable CpG methylation sites within ASMs in the mouse genome will be important (Barbaux et al., 2012; Court et al., 2014; Stelzer et al., 2013; Wang et al., 2014; Xie et al., 2012). The boundaries for DNMT1-dependent methylation of known ASMs and other NORED regions were presented in Fig. 2.2, Table 2.1, and Supplemental Table S2.2. While the overall methylation at established gASM regions for both *Rasgrf1* and *H19* showed partial recovery in r1KO ESCs, both had smaller NORED regions, within the larger established gASMs, that were completely lost (and not rescued) in DNMT1-deficient cells (Fig. 2.2A, 2.2I). Overall, this suggests two possible scenarios: 1) that only a subset of CpG sites within germline imprinted loci are NORED or 2) that NORED accurately defines CpG sites responsible for regulating parent-of-origin allelic methylation at imprinted loci. We favor the latter because it is consistent with the mechanistic understanding that DNMT1 is absolutely critical for maintaining parent-of-origin allelic DNA methylation at germline imprints (Biniszkiewicz et al., 2002; Hirasawa et al., 2008).

A mosaic is comprised of smaller subunits from which a larger pattern emerges. Analogously, our 'MethylMosaic' approach reveals DNA methylation patterns at a genomic region that is composed from whole genome bisulfite sequencing reads, based on the principles that each sequencing read represents a separate DNA molecule (see Methods). By evaluating read-level methylation, as opposed to CpG-level methylation, we observe emergent patterns at the molecular level that can be used to provide mechanistic insight. While many patterns could be explored, highly methylated and lowly methylated reads–which are quantified as hyper- and hypomethylation indices–have been presented here to detect genomic regions with bimodal methylation that occurs at ASMs (Fig. 2.3 and Supplemental Table S2.3).

By combining DNMT1-dependent DNA methylation loss with the bimodal methylation patterns characteristic of allelic methylation, we have presented regions that share epigenetic properties with gASMs. Unexpectedly, we identified additional regions (e.g., *Park7* and *Hcn2*) that although both vulnerable and bimodal, do not display parent-of-origin dependent allelic methylation in reciprocal cross ESCs. While further detailed characterization of novel bimodal NORED regions is needed to examine parental and genetic influences on allelic methylation at other loci, we have validated parent-of-origin independent bimodal methylation of two loci, *Park7* and *Hcn2*.

Cross talk between genetics and epigenetics via non-imprinted ASMs in the mammalian genome

Genetic variations have long been associated with common diseases. Over the past decades, numerous genome-wide association studies (GWAS) have been performed for many common diseases, including diabetes, autoimmune diseases, and neurological disorders (Simón-Sánchez and Singleton, 2008). Allele frequencies of hundreds of common variants are reported as statistically correlated with diseases. However, it is controversial whether these variants have biological relevance to disease pathogenesis and clinical prognosis or treatment (McClellan and King, 2010; Visscher et al., 2012).

Contrary to early expectations of SNP variants disrupting protein-coding genes, the vast majority (about 88%) of GWAS-identified SNPs reside in intergenic or intronic regions, which puzzle the community for understanding their functional relevance. Herein, our investigations provide a clue that a single nucleotide difference (rs240718423 in Cast allele; Fig. 2.4C) could have functional relevance. This SNP variant may predispose Cast allele to be hypomethylated, whereas the other allele hypermethylated at *Hcn2*/*Polrmt* ASM. While further investigation is needed for demonstrating the influence of *Hcn2*/*Polrmt* ASM on the expression of *Hcn2* and *Polrmt*, published data do confirm that the region surrounding this ASM is associated with inactive marks including H3K9me3, H4K20me3, and H3K27me3 and *Hcn2* promoter is associated with inactive H3K27me3 (see Fig. 6 in ref. (Mikkelsen et al., 2007)). These inactive marks are widely considered as inactive histone marks for repressing gene transcription (Barski et al., 2007; Dai and Wang, 2014; Wang et al., 2008). Because *Hcn2* is highly expressed in ESCs, these inactive marks are expected to be associated with the silent allele (presumably hypermethylated 129 and C57 allele, not the hypomethylated Cast allele). Note, the coverage of the published ChIP-seq data was not enough to call

allelic chromatin marks (Mikkelsen et al., 2007). Altogether, our results combined with previously reported findings reveal a possibility that the difference at one single nucleotide could be amplified through alteration of local chromatin structure, thereby changing the fate of a gene on a given allele.

*Regional **a**utosomal **c**hromosome **i**nactivation (ACI): an X-chromosome inactivation (XCI)-like mechanism in controlling autosomal genes in mammals?*

Not all genes on the inactivated X chromosome are silenced. In other words, it is a mechanism of regional inactivation of X chromosome. The selection of inactivated X chromosome in eutherian (placental) mammals, such as mice and humans, involves the transcription of a master regulator *Xist* (X-inactive specific transcript), a long non-coding RNA, as well as the expression of antisense *Tsix* of *Xist*. DNA methylation and inactive histone modification marks are necessary for the inactivation of selected X chromosome.

The feature above resembles that of *Hcn2*/*Polrmt* ASM and other ASMs. Methylation of CpG sites happens at one allele (e.g., C57 or 129), but not the other Cast allele at the *Hcn2*/*Polrmt* locus. Inactive histone modification marks including H3K9me3, H4K20me3, and H3K27me3 occupy the ASM and regions beyond the ASM (Mikkelsen et al., 2007). Given the fact that *Hcn2* is highly expressed in ESCs (in report (Mikkelsen et al., 2007) and our own data) and that these three inactive marks are associated with silent genes (Barski et al., 2007; Wang et al., 2009), it is reasonable to expect that C57 or 129 allele (when combined with Cast allele) was the repressed allele. In addition, our unpublished observation suggests that *Hcn2*/*Polrmt* ASM- (and other ASMs) derived repressive chromatin structure extend beyond ASM region. In other words, the ASM-derived repressive chromatin controls the expression of genes in a long genomic region, resembling the inhibition of genes by repressive chromatin on inactivated X chromosome. The expected switchable feature of some ASMs (at least *Hcn2*/*Polrmt* ASM in J1 and H1 ESC lines; *Park7* ASM in four tested mouse ESC lines) resembles the observation of random inactivation of one of the two X chromosomes. Lastly, *Hcn2*/*Polrmt* locus also shares the feature of having a long non-coding RNA, BC1. Whether BC1 plays similar roles to *Xist/Tsix* in the case of inactivation of one X chromosome or to *Kcnq1ot1* in the case of repression of non-imprinted allele at *Kcnq1*/*Kcnq1ot1* locus, however, requires further investigation. All together, these shared features raise the possibility that a

regional autosomal chromosome inactivation (ACI), like XCI mechanism, is controlling some autosomal genes. We envision that this ACI acts regionally (i.e., controlling only a subset of genes on a given autosome; not entire chromosome) and that these autosomal genes would be centered around the identified bimodal MethylMosaic ASMs. A close concept of "parallel to XCI" is also recently proposed from Dr. Edith Heard group (Gendrel et al., 2016).

4. Materials and Methods

Cell culture, DNMT1-rescued 1KO cells, and human ESC differentiation

Mouse ES cells were cultured as described before (Li et al., 2015). Briefly, mouse ES cells (J1, 1KO, DKO, TKO) were maintained without feeder cells on 0.1% gelatin coated Petri dish in DMEM medium supplemented with 15% FBS (ES cell grade), 2mM glutamine, 10uM mercapto-ethanol, 100U/ml LIF, Penicillin/Streptomycin mixture 100ug/ml, 1X non-essential amino acid. Cultured DNMT1 KO (1KO) cells were transfected with constructs expressing GFP-fused DNMT1 (kind gift of Dr. Heinrich Leonhardt (Spada et al., 2007)). DNMT1-GFP-rescued 1KO cells were sorted using facility of Bloomberg School of Public Health.

Four hybrid ES cell lines were directly derived from the blastocysts, which were from the reciprocal crosses between mice on different genetic backgrounds (e.g., Cross between Cast/EiJ and 129S1/SvimJ or cross between Cast/EiJ and C57BL/6J). Primary mouse embryo fibroblasts inactivated by Mitomycin C were used as feeder cells. ES Cells were expanded on pre-coated plates with 0.1% Gelatin in LIF (+) medium on feeder cells and then moved to feeder-free 2i medium (EMD Millipore) to get rid of feeder cells.

The differentiation of human H1 ESCs (Wi Cell, Madison, WI) into neuronal progenitor cells (NPCs) and neurons were performed similarly using our developed RONA (rosette-type neural aggregates) method (Martin et al., 2014; Xu et al., 2016). Briefly, detached hESC colonies were grown in suspension in human ES cell medium without FGF2 (defined as knockout serum replacement medium) in low attachment 6-well plates (Corning, Corning, NY), supplemented with Noggin (50ng/ml; R&D systems, Minneapolis,

MN) or Dorsomorphin (1μM, Tocris Bioscience, Bristol, UK) and SB431542 (10μM, Tocris Bioscience) from day 2 to day 6. Free-floating embryoid bodies (EBs) were attached and supplied with N2-induction medium (NIM) containing DMEM/F12 (Invitrogen, Carlsbad, CA), 1% N2 supplement (Invitrogen), 100 μm NEAA (Invitrogen), 1 mM Glutamax (Invitrogen), and heparin (2 μg/ml; Sigma, St. Louis, MO) from day 7 to day 16. Highly compact 3-dimensional column-like neural aggregates were collected and maintained as neurospheres in Neurobasal medium containing B27 minus vitamin A (Invitrogen), 1 mM Glutamax 1 day. For neuronal differentiation, dissociated neurospheres were maintained in neural differentiation medium containing Neurobasal/B27 (NB/B27, Invitrogen), BDNF (20ng/ml, PeproTech, Rocky Hill, NJ), GDNF (20 ng/ml, Peprotech), ascorbic acid (0.2 mM, Sigma), and dibutyryl cAMP (0.5mM, Sigma).

Bisulfite whole genome sequencing (BS-seq) library construction

One to five micrograms of genomic DNA of DNMT1-rescued 1KO cells was fragmentized to 200~500bp by Covaris S2 sonicator (Covaris, USA). End repairing was then performed following manufacturer's instruction (End-It DNA end repair kit, Epicentre, USA). After Ampure XP (Sigma USA) purification, adenine was added to 3' end with 3ml DNA Taq polymerase (M0267S, NEB) and 1mM dATP in 50ml reaction solution incubated at 70°C for 30 mins. After Ampure XP purification, 1ml of Illumina Trueseq adaptors were ligated with 4ml T4 DNA ligase (M0202L, NEB) in 40ml reaction solution and incubated at 16°C overnight. Fragments at 300-600 bp from adaptor-ligated DNA were collected from 2% agarose gel, and then bisulfite-treated using Imprint DNA modification Kit (MOD50-1 KIT, Sigma USA) as manufacturer instructed. PCR enrichment was performed to amplify the libraries, which were then collected from 2% agarose gel electrophoresis at size 300-600bp. Hi-seq 2000 was used for generating all deep-sequencing data.

Data processing of BS-seq data

Trim Galore version 0.4.0 using Cutadapt version 1.8.1 and FastQC version 0.11.2 was used to trim sequencing reads (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Bismark version 0.14.5 (Krueger and Andrews, 2011) implementing Bowtie2 version 2.2.5 (Langmead and Salzberg, 2012)

was used to align trimmed reads to using options -N 1 -D 20 -R 3 -X 1000 --chunkmbs 1024. To generate a strain-specific reference genome for J1 and J1-derived ESCs, we substituted mouse strain 129S1 SNPs from Mouse Genomes Project (Keane et al., 2011) SNP Release version 5 (REL-1505-SNPs_Indels) into GRCm38/mm10 genome. Bismark deduplication was used to remove PCR duplicates from aligned pairs and Bismark methylation extractor was used to determine the methylation status of cytosines To prevent methylation bias at the ends of reads, we removed methylation calls in the first eight base pairs of each read (Krueger and Andrews, 2011). R version 3.2.2 was used for post-processing analyses. We merged CpG methylation calls on positive and negative strands into single, destranded CpG sites.

Methylation of 21 well-characterized germline imprinted ASMs (gASMs)

Well-characterized gASMs were defined as the 21 gASMs identified in common between two sources (Arnaud, 2010; MacDonald and Mann, 2013) (Supplemental Table S2.1). Methylation across a gASM was calculated as the average of all covered CpG sites. Comparison of methylation in WT, 1KO, DKO, and TKO at these loci was determined by pairwise Wilcoxon rank-sum tests (paired) and p-values were adjusted for multiple comparisons by Bonferroni method.

No Restored DMRs (NORED)

The concept for NORED is that in order for a CpG site to be considered not restorable in DNMT1-rescued 1KO cells it would need to meet three criteria: 1) that it has sufficient methylation in WT, 2) that it experiences near complete methylation loss in 1KO, and 3) that it recovers minimum methylation in r1KO. To determine methylation levels in WT, 1KO, and r1KO ESCs to use as cutoffs for these criteria, we performed an ROC-like analysis using a permutation-based approach. Only CpG sites from autosomal chromosomes with at least 5x coverage in each cell type (WT, 1KO, and r1KO ESCs) were used. To estimate the false positives at various combinations of methylation cutoffs for WT, 1KO, and r1KO, we simulated twenty null datasets by swapping the cell type labels at each CpG site among the three ESCs. For WT, we tested 0.25 and 0.30 as minimum values. For 1KO, we tested 0.03 and 0.05 as maximum values. For combinations of WT and 1KO, we tested r1KO between WT and 1KO at increments of 0.005 for maximum values. We applied these criteria to null datasets and considered all positives identified as false

positives and all negatives as true negatives. FPR for each null dataset was calculated as the number of false positives divided by the sum of false positives and true negatives. FPR presented is the average FPR for twenty null datasets. We then applied the same cutoff combinations (described above) to the original data to identify positives and negatives. At FPR of 0.01, 70 percent of CpG sites within gASMs (defined as those in common from two sources (Arnaud, 2010; MacDonald and Mann, 2013)) had at least 25 percent methylation in WT, at most 5 percent methylation in 1KO, and at most 12.5 percent methylation in r1KO; genome-wide 3.3 percent of CpG sites met these criteria (Supplemental Fig. S2.2A). We therefore chose these values as the criteria to use to identify non-restored methylation at CpG sites.

To identify NORED regions, we developed a simple scoring system to combine nearby non-restored CpG sites into larger regions, allowing for an occasional restored CpG site to be included. At each CpG site, we assigned two points if it met the criteria and one point for each of the two proceeding and two following CpG sites that met the criteria, for a maximum point of six points per CpG position. We then clustered individual CpG sites in to regions with the regionFinder function from Bump Hunter (Jaffe et al., 2012), using a cutoff of three points for inclusion. Cumulative score was used to rank regions, thereby taking into account both the number of CpG sites and the proportion of CpG sites within the region that meet the criteria. To determine the FDR for NORED, we applied the scoring and clustering algorithms to the permutated datasets to identify false positive regions. For each unique cumulative score in the original data, we calculated the number of regions considered positive at this threshold in the real data and in each null dataset. We then divided the number of false positives by the number of positives to calculate the FDR for each permuted dataset. The average FDR of twenty null datasets is presented as the FDR for the method.

MethylMosaic

In diploid organisms, gASM is characterized by one allele being methylated and the other allele being unmethylated. Conventional sequencing of PCR clones to confirm gASM prompted us to develop MethylMosaic. It is expected that roughly half of sequencing reads (or tags) from these regions carry high proportion of methylated CpG sites, while the other half carry high proportion of unmethylated sites, leading to a bimodal distribution of methylation at these regions. For regions without gASM, all sequencing

reads should have similar proportions of methylated CpG sites since the two parental chromosomes would have equivalent methylation level at these loci.

To calculate read-level methylation for WT ESCs, the methylation of each CpG site within a sequencing read was determined by comparing read sequence with reference genome where a conversion from Cytosine on reference genome to Thymine on read sequence indicates unmethylated status and no such conversion indicates methylation. The overall methylation of a read is calculated by dividing the number of methylated CpG sites by the total CpG sites covered by the read.

Hypomethylation and hypermethylation indices were introduced to identify bimodal regions. For each CpG site on autosomal chromosomes in the mouse genome, a window enclosing 300 bp upstream and 300 bp downstream of that site was defined. Sequencing reads overlapping each window were retrieved and hypomethylation index was calculated as the proportion of reads with at most ten percent methylation. Similarly, hypermethylation index was calculated as the proportion of reads with at least ninety percent methylation. To select values of hypo- and hypermethylation indices to use as criteria to consider a CpG site bimodal, we applied various cutoff ranges (combinations of upper and lower bounds) to both hypo- and hypermethylation indices. We calculated the TPR as the proportion of 21 well-characterized gASMs, defined as those in common from two sources (Arnaud, 2010; MacDonald and Mann, 2013), identified for each cutoff range. To calculate the FPR, we simulated ten null datasets by shuffling the methylation calls at each CpG site among the reads that cover it. Notably, the randomization of methylation calls among reads at each CpG site has the potential to alter read-level methylation only. That is, the methylation at a given CpG site remains the same as it was prior to randomization. We calculated hypo- and hypermethylation indices and applied cutoff ranges (as described above) to identify the positive CpG sites within the null datasets (as described above). Because all positive CpG sites in the randomized data sets were considered false positives, we calculated the FPR for each cutoff range as the number of identified CpG divided by the number of CpG sites in the genome. The FPR was determined for each cutoff range by averaging the FPR from ten null datasets. We selected 0.2 for the lower bound and 0.75 for the upper bound as criteria for both hypermethylation and hypomethylation indices based on the ROC curve (Supplemental Fig. S2.3A). That is, we considered a CpG site to be bimodal if the proportion of hypomethylated reads within the defined

window was at least 0.2 and at most 0.75 and the proportion of hypermethylated reads within the defined window was at least 0.2 and at most 0.75.

To cluster individual CpG sites into regions, we combined consecutive bimodal CpG sites. For WT data, we ranked bimodal regions by the how many CpG sites were included. For each unique rank (ties were assigned to lowest rank), we determined the number of regions that would exceed the threshold for quantity of CpG sites within the region (i.e., number of regions considered positive at each rank). For each of the randomized data sets we combined individual false positive CpG sites into regions and calculated false positive bimodal regions as the number of regions exceeding the CpG site quantity threshold for a given rank of WT data. To calculate the FDR for each rank, we divided the number of false positive bimodal regions by the number of regions identified as positive in the WT data. FDR was calculated for each null data set and averaged to determine the region-level FDR for bimodal regions by MethylMosaic method.

Enrichment Analyses for Chemical and Genetic Perturbations

MSigDB version 5.2 was used to identify overrepresented gene sets for Chemical and Genetic Perturbations from Curated Gene Sets at an FDR less than 0.05 (Subramanian et al., 2005). For each region set, all genes overlapping the center of the region, with a transcription start site within +/- 100 bp from the center of a region, or nearest the gene (if region is intergenic and does not overlap gene annotations) were considered to associated with a region.

Bisulfite Sanger Sequencing

DNA was extracted from four hybrid ESC lines, H1 ESCs, H1-differentiated NPC/neurons using phenol/chloroform/isoamyl alcohol (25:24:1). About 10 µg of those DNA was bisulfite converted and purified by Sigma Aldrich's Imprint® DNA Modification Kit following manufacturer's protocol. Exact 2µl of the bisulfite converted DNA was used as template for PCR amplification with KAPA HiFi HotStart Uracil Plus (Boston) used for the PCR reaction. For *Hcn2* ASM mouse amplification, the primer sequences were as follows: forward, GGT GTA GTA GGT AGA GTT TGG TTA G and reverse, CTC AAA AAT CAC AAA TTA AAA AAA A were used to amplify a 529 bp. For *Park7* ASM, the primer sequences were

as follows: forward, TTT AGG TGA ATT TTT GGA ATT GTT T and reverse, CCT TCC CTA ACT ACT TAA ATT AAC AC were used to amplify a 334bp. Amplicons were ligated to PMD19 vectors (Clonetech) with T4 DNA ligase (M0202L, NEB) and transformed into DH-5α competent cells (NEB). Monoclonal bacteria colonies carrying the plasmid were cultured and picked up from plates with AMP and X-gal. All the colonies were sequenced at the Genewiz (Boston).

For H1, H1-derived NPCs, and H1-derived neurons, primers based on conserved sequence of mouse genome: forward, GAG ATG GTG TAG TAG GT and reverse, ACC AAA TAT TAC ACT TAA AAA A were used to amplify a 977 bp fragment inside of the potential DMR in the human genome for HCN2 gene. Amplicons were ligated to TA vectors from PCR cloning Kit (Invitrogen) and transformed into DH-5α competent cells. Monoclonal bacteria colonies carrying the plasmid were cultured and picked up from plates with AMP and X-gal. All the colonies were sequenced at the Beckman Genomic Com (Danvers, Massachusetts). Sequenced reads were aligned back to the Bisulfite converted genome sequences by Seqman pro (Lasergene). PCR replicates were discarded and methylation for monoclonal reads was methylation was determined with BiQ analyzer (Bock et al., 2005).

5. List of Supplemental Materials

Supplemental Figures S2.1 to S2.2 (**Appendix I**)

Supplemental Tables S2.1 to S2.8

Chapter 3

Regional DNA methylation characteristics influence susceptibility to epigenetic perturbation from prenatal exposure to Bisphenol A in male fetal livers

Abstract

Epigenetic modifications, such as DNA methylation, regulate genomic processes and serve at the interface between genes and the environment. Both cell- and tissue-specific regulation of epigenetic mechanisms contribute to maintenance of physiological homeostasis. Life stage-specific regulation of epigenetic processes is necessary for proper development. Nutritional and environmental exposures that alter epigenetic marks can contribute to disease pathophysiology or lead to increased risk for diseases at later life-stages. Imprinted genes and transposable elements represent two unique paradigms of epigenetic inheritance. As such, determining how the epigenetic marks at these genomic features are affected by exposure to environmental pollutants will provide insight into mechanisms of toxicity during critical periods of development. We previously reported a division of labor between DNA methyltransferase (DNMT) enzymes, DNMT1 and DNMT3a/3b, to suppress distinct types of retrotransposons. Additionally, DNMT1-dependent loss of methylation at known imprinted DMRs could not rescued by exogenous *Dnmt1* cDNA expression. To determine whether DNMT1-dependent genomic regions are more susceptible to environmental perturbation, we compared fetal liver methylomes from male C57BL/6J mice that were exposed *in utero* to Bisphenol A (BPA) through maternal supplemented diet (0 μg/kg bw/d, 10 μg/kg bw/d, or 10 mg/kg bw/d). We found that DNMT1-maintained and non-restorable genomic regions were more susceptible to environmental perturbation and had altered CpG methylation in response to BPA exposure.

1. Introduction

Early life exposure to nutritional and environmental stressors can influence phenotypes over multiple generations. At certain developmental stages during the mammalian lifecycle, up to three generations can be simultaneously exposed. That is, maternal exposures during the gestation period can affect the mother (F0), the developing fetus (F1), and the germ cells of the developing offspring (F2). Intergenerational inheritance refers to the molecular changes and phenotypes that result from direct exposure to multiple generations. Transgenerational inheritance occurs when molecular changes and phenotypes persist in subsequent generations (e.g., F3) that were not directly exposed to the stressor. Epigenetic modifications, such as DNA methylation, have been proposed as possible molecular mediators for non-genetic inheritance of environment-induced phenotypes at later life stages and/or across

generations. DNA methylation dynamics over the course of the mammalian lifecycle, specifically genome-wide methylation reprogramming following fertilization and during germ cell specification, serve as a barrier for widespread epigenetic inheritance across generations. Nonetheless, parent-of-origin dependent methylation/expression of imprinted genes and transgenerational silencing of transposable elements are paradigms of epigenetic inheritance.

Animal models used to explore mechanisms for environment-induced epigenetic changes from perinatal exposures have reported altered DNA methylation at several imprinted genes (Iqbal et al., 2015; Kang et al., 2014; Susiarjo et al., 2015) and metastable epialleles (Anderson et al., 2012; Dolinoy et al., 2007b; Faulk et al., 2014). In the viable yellow agouti mouse model, the A^{vy} mutation comes from an Intracisternal A Particle (IAP) element, a type of long terminal repeat (LTR) retrotransposon, insertion upstream of an endogenous gene that regulates coat color. DNA methylation level at the IAP insertion correlates with coat color in A^{vy}/a heterozygotes and is sensitive to environmental pollutants, including Bisphenol A (BPA), and nutritional factors (Dolinoy et al., 2006; 2007b). Metastable epialleles with observable phenotypes (e.g., A^{vy} and coat color) are used as biosensors for epigenotoxicity because of their vulnerability to epigenetic perturbation (Dolinoy, 2008). In addition to DNA methylation at metastable epialleles, DNA methylation at other repetitive element loci are used as an indicator of global methylation levels (Yang et al., 2004).

In mammals, three DNA methyltransferase (DNMT) enzymes are responsible for adding methyl groups to DNA. DNMT1 is classified as the "maintenance" enzyme because has a preference for hemimethylated DNA *in vitro* (Bestor and Ingram, 1983; Gruenbaum et al., 1982) and localizes to replication foci to copy methylation information from parental strands to daughter strands following DNA replication (Leonhardt et al., 1992). DNMT3a and DNMT3b are considered *"de novo"* methyltranferases due to their role in establishing new methylation patterns in during embryonic development (Okano et al., 1999). Once established by DNMT3a/3b in early embryos, methylation patterns will be "faithfully" transmitted in subsequent divisions of somatic cells. However, the aforementioned, traditional two-step model (i.e., DNMT3a/3b for establishment and DNMT1 for maintenance) does not explain the diversity in DNA methylation changes observed throughout the genome in response to environmental or genetic

influences. We have previously used mouse embryonic stem cells (mESCs) to show the extent to which genomic features depend on DNMT1 or DNMT3a/3b activity to maintain DNA methylation patterns. Most genomic features lost substantial methylation in the absence of either DNMT1 or DNMT3a/3b, thereby demonstrating that many genomic features rely on activity from both "*de* novo" and "maintenance" DNMT enzymes for maintaining full DNA methylation (Li et al., 2015). However, there appears to be division of labor between DNMT1 and DNMT3a/3b to suppress distinct types of retrotransposons in mouse embryonic stem cells (mESCs). Specifically, long terminal repeats (LTRs) tend to be maintained by DNMT1, whereas, LINEs were maintained by DNMT3a/3b. Our findings lead to an updated "coordination and division of labor model" for DNA methylation maintenance. Notably, IAP elements were predominantly DNMT1 dependent, which suggests the DNA methylation maintenance paradigm at these loci influences their susceptibility to epigenetic perturbation from developmental exposures.

DNA methylation is dramatically reduced at germline imprinted loci in the absence of either DNMT1 or DNMT3a/b and methylation at these loci is not restored by expression of exogenous *Dnmt1* cDNA in *Dnmt1*$^{-/-}$ mESCs (Chapter 2; (Tucker et al., 1996)). Whole genome bisulfite sequencing (WGBS) revealed approximately 2500 genomic regions that display this non-rescued DMR (NORED) characteristic in mESCs (Chapter 2). Distinct from most of the genome, the paternally and maternally inherited alleles at imprinted germline differentially methylated regions (gDMRs) have different methylation patterns (i.e., one allele fully methylated and the other unmethylated). The parent-of-origin-dependent allelic methylation at gDMRs arises from differential methylation patterns that had been established in male and female germ cells. Such allelic methylation at gDMRs is resistant to global demethylation during post-fertilization epigenetic reprogramming phase. The property of one-methylated-allele and the-other-unmethylated-allele emerges as bimodal methylation patterns (i.e., half of DNA molecules at a locus are methylated and the other half are unmethylated), which can be observed with targeted or genome-wide bisulfite sequencing methods. Analyses of WGBS data revealed approximately 2500 genomic regions with bimodal methylation patterns in mESCs. Approximately 200 NORED regions had bimodal methylation patterns, referred to as "bimodal NORED" regions. Bimodal NORED regions were comprised of known imprinted germline DMRs and additional loci of unknown imprint status.

The unique DNA methylation characteristics observed at germline imprints and retrotransposon families, raises the possibility that the epigenetic properties of these regions contribute their apparent susceptibility to exogenous factors. To examine the susceptibility of these features to epigenetic perturbation, we compared fetal liver methylomes from male C57BL/6J mice that were exposed *in utero* to physiologically relevant doses of BPA through maternal supplemented diet. Dose groups are referred to as "control" (0 μg/kg bw/d; n = 3), "lower" (10 μg/kg bw/d, n = 4), and "upper" (10 mg/kg bw/d, n = 4). We used previously identified genomic regions — hereinafter referred to as "bimodal exclusive", "bimodal NORED", "NORED exclusive", "DNMT1-maintained", and "DNMT3a/3b-maintained" — based on DNA methylation features exhibited in mESCs. We found that CpG sites within these genomic regions had higher propensity for altered methylation from prenatal exposure to BPA than CpG sites genome-wide.

2. Results

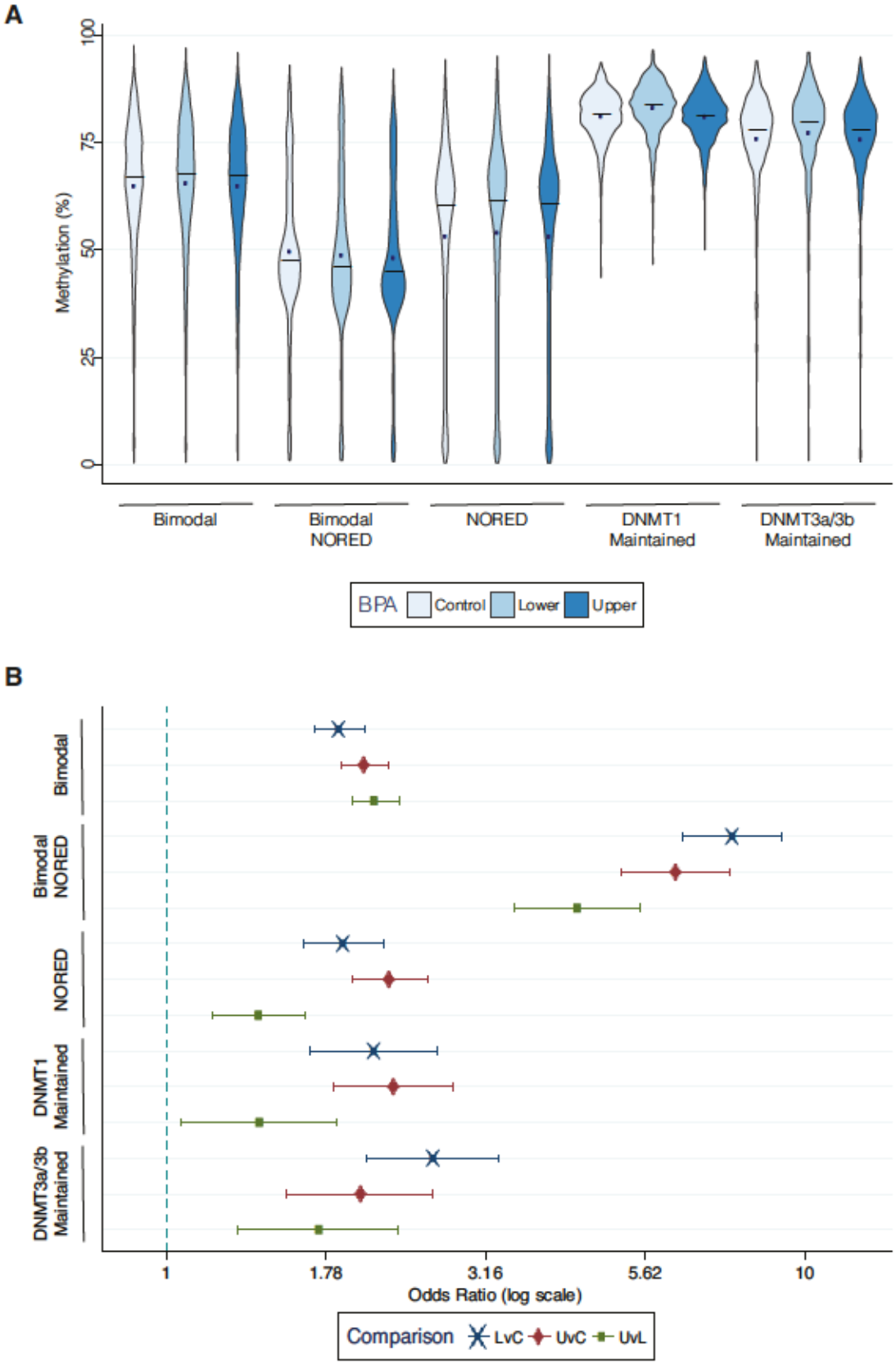
Propensity for altered methylation at CpG-sites in genomic regions with distinct DNA methylation characteristics in mESCs

We examined DNA methylation in control livers and in livers from male mice fetuses that had been prenatally exposed to BPA. We focused on five classes of genomic "regions of interest" that had been previously identified in mESCs based on either 1) shared epigenetic characteristics observed at imprinted gDMRs or 2) distinct DNA methylation maintenance paradigms. Regions with imprint-like characteristics can be classified as exclusively bimodal, bimodal NORED, or exclusively NORED. DNMT1-maintained and DNMT3a/3b-maintained loci are considered to have distinct DNA methylation maintenance paradigms in contrast with the majority of CpG sites genome-wide, which required DNMT1 and DNMT3a/3b enzymes for full methylation in mESCs (Li et al., 2015). That is, DNMT1-maintained regions were resistant to methylation loss in the absence of DNMT3a/3b and DNMT3a/3b-maintained regions were resistant to methylation loss in the absence of DNMT1. Within each region type, overall CpG methylation profiles were similar among control, lower and upper groups (Fig. 3.1A and Supplemental Fig. S3.1–S3.5). Average CpG methylation percent for bimodal, bimodal NORED, and NORED regions ranged from 64.9 to 65.5, 48.2 to 49.6, and 53.1 to 54, respectively. DNMT1-maintenance and DNMT3a/3b-maintenance regions were highly methylated in all dose groups (Fig. 3.1A and Supplemental Fig. S3.4–S3.5). With the

exception of bimodal NORED regions, methylation was slightly higher in the lower dose group relative to both control and upper dose groups. In bimodal NORED regions, average CpG site methylation decreased as dose increased.

To determine whether regions displaying distinct DNA methylation characteristics observed in mESCs were more susceptible to perturbation from prenatal BPA exposure, we used WGBS data from control, lower, and upper dose groups to identify CpG sites with altered methylation genome-wide. We defined the altered methylation as being within the lower 0.5^{th} percentile (hypomethylation) or the upper 99.5^{th} percentile (hypermethylation) of the test statistic for each pairwise comparison. To quantify the magnitude of association between CpG sites having altered methylation and residing within a region of interest, we use odds ratios (ORs). ORs are a commonly used approach for identifying an association between two variables (Edwards, 1963; Mosteller, 1968); Pubmed (NCBI Resource Coordinators, 2016) search for "odds ratio" returns over 200,000 results. As opposed to simply determining whether BPA altered methylation at CpG sites within our regions of interest, here we used ORs because we wanted to determine whether altered methylation was more likely to occur at CpG sites within specific genomic regions than CpG sites outside. For each region class (e.g., bimodal), the odds of a CpG site having altered methylation and being within the region class was divided by the odds of a CpG site having altered methylation and not being within the region class. ORs were calculated for lower versus control (LvC), upper versus control (UvC), and upper versus lower (UvL) comparisons.

Figure 3.1 Prenatal BPA exposure altered CpG methylation within bimodal, bimodal NORED, NORED, DNMT1-maintained, and DNMT3a/3b-maintained genomic regions. (A) Violin plots of methylation profiles for control (n = 3), lower (n = 4), and upper (n = 4) dose groups from male fetal livers following *in utero* exposure to BPA. Bar (—) indicates median; point (■) indicates arithmetic mean. **(B)** Odds ratios for altered methylation at CpG sites within each region compared to CpG sites outside of each region for lower versus control (LvC), upper versus control (UvC), and upper versus lower (UvL) comparisons. Error bars represent 95 percent confidence intervals corrected for multiple tests (15) via Bonferroni method for alpha < 0.05.


A
Methylation (%)
0
25
50
75
100
Bimodal
Bimodal NORED
NORED
DNMT1 Maintained
DNMT3a/3b Maintained
BPA
Control
Lower
Upper
B
Bimodal
Bimodal NORED
NORED
DNMT1 Maintained
DNMT3a/3b Maintained
1
1.78
3.16
5.62
10
Odds Ratio (log scale)
Comparison
LvC
UvC
UvL

For all comparisons, all region classes exhibited significantly higher odds of having altered CpG sites, than genome-wide CpG sites not within the region class (Fig. 3.1B). P-values were adjusted (adj p) for multiple tests (15) using Bonferroni method. For bimodal exclusive regions, odds ratios (ORs) ranged from 1.87 to 2.12 (LvC: adj p = 1.21×10^{-74}, UvC: adj p = 1.61×10^{-105}, UvL: adj p = 4.68×10^{-119}). For bimodal NORED regions, ORs ranged from 4.42 to 7.73 (LvC: adj p = 2.95×10^{-146}, UvC: adj p = 1.30×10^{-104}, UvL: adj p = 4.68×10^{-119}). For NORED exclusive regions, odds ratios ORs ranged from 1.40 to 2.24 (LvC: adj p = 4.39×10^{-31}, UvC: adj p = 3.35×10^{-55}, UvL: adj p = 2.34×10^{-07}). For DNMT1-maintained regions, odds ratios ORs ranged from 1.40 to 2.28 (LvC: adj p = 1.06×10^{-16}, UvC: adj p = 8.28×10^{-22}, UvL: adj p = 9.39×10^{-03}). For DNMT3a/3b-maintained regions, odds ratios ORs ranged from 1.74 to 2.63 (LvC: adj p = 6.01×10^{-24}, UvC: adj p = 1.92×10^{-11}, UvL: adj p = 2.51×10^{-06}).

Strikingly, bimodal NORED regions were most likely to have CpG sites with altered methylation. Since known imprinted gDMRs are included among the bimodal NORED regions, we wondered whether altered CpG sites were limited to known gDMRs. Examination of altered CpG sites within bimodal NORED regions demonstrates that altered methylation occurred in bimodal NORED regions not known to be imprinted, as well as in bimodal NORED regions within or near known imprinted gDMRs and (Table 3.1). While the overall methylation profiles for bimodal NORED regions showed lower methylation with increasing dose (Fig. 3.1A), hypermethylated and hypomethylation CpG sites occurred within these regions. Overall, CpG sites tended to change in a consistent direction within a locus, but several imprint-associated loci displayed both localized hypermethylation and hypomethylation within the larger imprinted region.

We next examined the direction of change for CpG sites with altered methylation in DNMT1-maintained and DNMT3a/3b-maintained regions. Altered CpG sites in DNMT1-maintained regions tended to be hypermethylated in lower dose compared to control dose (LvC: 65% of altered CpG sites were hypermethylated), but hypomethylated in upper dose compared to either control (UvC: 55% of altered CpG sites were hypomethylated) or lower (UvL: 59% of altered CpG sites were hypomethylated) dose. Interestingly, the majority of altered CpG sites within DNMT3a/3b-maintained regions were hypermethylated in all comparisons (LvC: 56%; UvC: 56%; UvL: 81%).

Table 3.1 Bimodal NORED regions with altered CpG site methylation

Locus	Lower vs. Control	Upper vs. Control	Upper vs. Lower
Gtl2 (*Meg3*) ICR*[a]	↑↓	↓	↑↓
Gtl2 (*Meg*3)[b]	↑↓	–	↑
H19 ICR[a]	↓	–	–
H19 promoter[b]	↑	–	↓
Igf2r ICR[a]	–	↑	–
Kcnq1ot1 ICR[a]	↑↓	↑	↑
Kcnq1/*Kcnq1ot1*[b]	↑	↑	–
*Nespas-Gnas*XL ICR[a]	↓	↓	–
Peg1 (*Mest*)[a]	↑↓	↑	↑
Peg10[a]	↓	↓	↓
Peg10/Sgce promoter[b]	↓	–	–
Peg13/*Trappc9*[a]	↑	–	–
Peg13/*Trappc9*[b]	↑	–	–
Rasgrf1 ICR[a]	↑	–	↑
Rasgrf[b]	–	–	↑
Snrpn ICR[a]	↑	↑	↑↓
Zrsr1/*Commd1* (*Murr1*)[a]	↓	↑	↑
Zrsr1/*Commd1* (*Murr1*)[b]	↓	–	–
H13/*Mcts2*[b]	–	–	↓
Ddx39b/*H2-L*	↓	–	↑
Fhad1	↓	–	–
Gm13212	–	↑	–
Hcn2/*Bc1*/*Polrmt*	–	↑	–
Lmna	↑	–	–
Mt3	–	–	↓
Nat8l	↑	↑	–
Nupr1l	↓	–	–
Pigk	–	↑	–
Rusc2	↓	–	–
Tbcd/*Zfp750*	↑	–	–
Tmem40	↑	–	↓
Trim80	–	↓	↓
4930481A15Rik	↓	–	↑
chr6:127251339-127251463	↓	–	–
chr10:79322231-79322387	↑	–	–
chr11:20446254-20446353	↑	–	–

* Imprint Control Region

[a] Altered CpG site(s) within known germline differentially methylated region

[b] Altered CpG site(s) near, but not within known germline differentially methylated region

↑ Hypermethylation

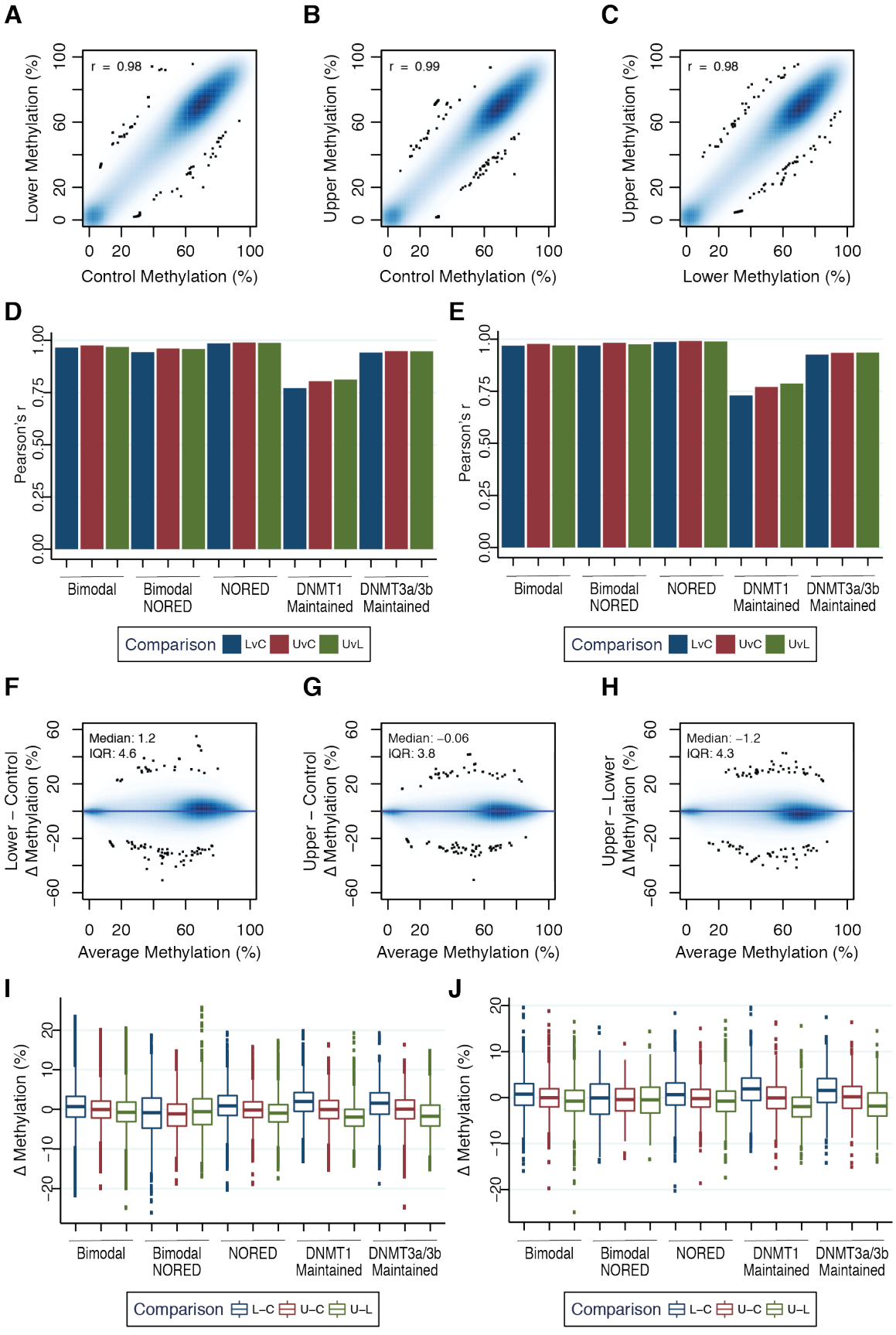
↓ Hypomethylation

↑↓ Small regions with either hypermethylation or hypomethylation within a larger genomic region

Prenatal BPA exposure induced genome-wide and regional shifts in methylation levels

Genome-wide, methylation of CpG sites was strongly correlated between control, lower, and upper exposure groups (Fig. 3.2A–C). The Pearson's r correlation coefficient for lower compared to control or upper was 0.98 ($p < 1 \times 10^{-13}$) and the Pearson's r correlation coefficient for upper compared to control was 0.99 (N = 20,041,855 CpG sites, $p < 1 \times 10^{-13}$). At the CpG site level, correlation among dose groups was high for bimodal (N = 57,827 CpG sites), bimodal NORED (N = 4,014 CpG sites) and NORED (N = 22,314 CpG sites) regions (Fig. 3.2D). DNMT3a/3b-maintained regions (N = 5,876 CpG sites) had slightly weaker similarity and DNMT1-maintained regions (N = 7,913 CpG sites) displayed the weakest similarity between dose groups (Fig. 3.2D).

Figure 3.2 Characterization of genome-wide and regional differences in methylation in BPA-exposed and control fetal livers. (A–C) Correlation (Pearson's r) between methylation at CpG sites genome-wide in lower vs. control **(A)**, upper vs. control **(B)**, and upper vs. lower **(C)**. **(D–E)** Correlation coefficients for CpG site- **(D)** or region-level **(E)** methylation at five regions types and three comparisons: lower vs. control (LvC), upper vs. control (UvC), and upper vs. lower (UvL). **(F–H)** Bland-Altman plots showing genome-wide CpG site methylation differences (Δ) between lower and control **(F)**, upper and control **(G)**, and upper and lower **(H)** dose groups. Average methylation (x-axis) indicates the average methylation at a CpG site for two dose groups within each comparison. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated by subtracting the 75th percentile and the 25th percentile change in methylation values. **(I–J)** Box plots showing CpG site- **(I)** or region-level **(J)** methylation differences for five region types and three comparisons: lower minus control (L-C), upper minus control (U-C), and upper minus lower (U-L).


A
B
C
r = 0.98
r = 0.99
r = 0.98
Lower Methylation (%)
Upper Methylation (%)
Upper Methylation (%)
Control Methylation (%)
Control Methylation (%)
Lower Methylation (%)
D
E
Pearson's r
Bimodal
Bimodal NORED
NORED
DNMT1 Maintained
DNMT3a/3b Maintained
Comparison
LvC
UvC
UvL
F
G
H
Median: 1.2
IQR: 4.6
Median: −0.06
IQR: 3.8
Median: −1.2
IQR: 4.3
Lower – Control Δ Methylation (%)
Upper – Control Δ Methylation (%)
Upper – Lower Δ Methylation (%)
Average Methylation (%)
I
J
Δ Methylation (%)
L–C
U–C
U–L

To examine the region-level similarity in methylation between dose groups, we averaged the CpG site-level methylation across all CpG sites within each element of a region type. As such, each element of a region type has equal weight. Similar to the high correlation in methylation among dose groups observed at the CpG site level, bimodal (N = 2319 regions) and NORED (N = 2260 regions) regions displayed high correlation in region-level methylation, whereas bimodal NORED regions (N = 207 regions) had somewhat higher correlation at the region-level in the LvC comparison (Fig. 3.2E and Supplemental Fig. S3.1–S3.3). Both DNMT1-maintained (N = 808 regions) and DNMT3a/3b-maintained regions (N = 484 regions) had slightly lower correlations in region-level compared with CpG site-level methylation (Fig. 3.2E and Supplemental Fig. S3.4–S3.5). In contrast to genome-wide methylation, upper dose showed higher correlation with lower dose than with control in DNMT1-maintained regions (Fig. 3.2D–E).

To further illustrate that BPA affects genome-wide and regional methylation, we examined the distribution of changes in methylation between the dose groups. For change in methylation, we calculated the change in methylation for individual CpG sites and report the median and the interquartile range (IQR; i.e., difference between 75th percentile and the 25th percentile change in methylation values) for all CpG sites. IQR, which encompasses the middle 50 percent of the data, is an indicator of dispersion. We found a slight global hypermethylation in lower dose compared to control and upper dose. For LvC, median genome-wide change in methylation was 1.2 percent and the interquartile IQR was 4.6 percent (Fig. 3.2F). For UvC, median genome-wide change in methylation was -0.06 and the IQR was 3.8 percent (Fig. 3.2G). For UvL, median genome-wide change in methylation was -1.2 and IQR was 4.3 percent (Fig. 3.2H). Despite a similar direction and magnitude of change in global methylation in lower compared to either control or upper dose group, CpG site methylation change varied more between lower and control than between lower and upper dose groups.

We next characterized the distribution of CpG site-level methylation changes in our five regions of interest (Fig. 3.2I). For NORED and bimodal regions, median methylation change was less than one percent and followed the direction of the genome-wide shift toward slightly increased methylation in lower compared to control and upper dose groups. DNMT1-maintained and DNMT3a/3b-maintained regions also shifted toward increased methylation at lower dose (approximately two percent compared to either control

or upper), but had a larger shift in methylation than observed than genome-wide. Consistent with regional methylation profiles (Fig. 3.1A), median CpG site methylation decreased in bimodal NORED regions for all comparisons, with the largest shift toward hypomethylation observed in upper dose compared to control. For all comparisons, IQRs for difference in methylation among CpG sites were broadest for bimodal NORED regions (LvC IQR = 7.6%; UvC IQR = 5.5%; UvL IQR = 4.2 %) and greater than the genome-wide IQRs. This indicates that methylation at CpG sites in bimodal NORED regions exhibited higher variation in BPA-induced methylation changes than CpG sites in other regions. Other regions had IQRs less than 5.5 percent, with the highest variation occurring within the LvC comparison for each region. NORED and bimodal regions varied least in the UvC comparison. DNMT1-maintainted and DNMT3a/3b-maintained regions varied least in UvL comparison.

We next examined the region-level change in methylation for the five region types. For region-level change in methylation, we averaged the CpG site-level methylation across the CpG sites within each element of a region type and then calculated the change in methylation for each element. The median and IQR reported, therefore, represent the distribution of methylation differences for elements within each region type. At the region level, changes in methylation reflected those at CpG level (Fig. 3.2I–J and Supplemental Fig. S3.1–S3.5). That is, NORED, bimodal, DNMT1-maintained, and DNMT3a/3b-maintained regions followed genome-wide shift toward increased methylation at lower dose, whereas bimodal NORED regions were more likely to have decreased methylation at both lower and upper dose compared to control. Region-level IQRs for changes in methylation also reflected those at the CpG level (Fig. 3.2I–J and Supplemental Fig. S3.1–S3.5).

Characterization of methylation change at repetitive elements

Having previously identified LTR-type retrotransposons among regions preferentially maintained by DNMT1 (Li et al., 2015), we suspected that the low correlation between dose groups observed in DNMT1-maintained regions might be reflected in certain repetitive element families. With the above findings in mind, we next examined the region-level methylation for repetitive elements. Correlation was at least 0.84 ($p < 1 \times 10^{-13}$) for all repeat classes in all comparisons, which is higher than the correlation observed for DNMT1-maintained regions (Fig. 3.3A).

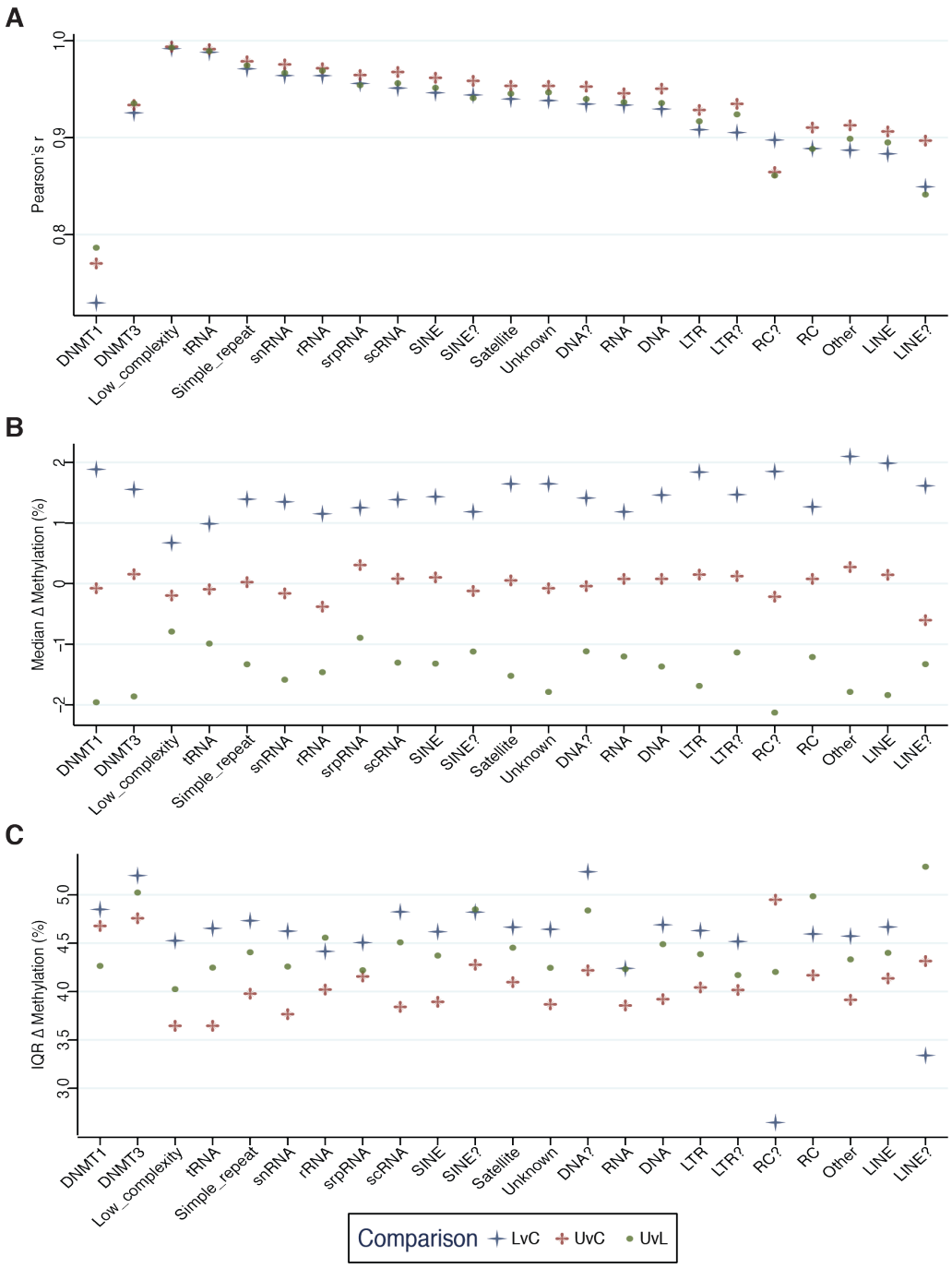


Figure 3.3 Characterization of region-level differences in methylation for repeat classes in BPA-exposed and control fetal livers. (A–C) Correlation coefficients **(A)**, median **(B)** and interquartile range (IQR) **(C)** for change (Δ) in methylation for three comparisons: lower vs. control (LvC), upper vs. control (UvC), and upper vs. lower (UvL). IQR measures the dispersion of the middle 50 percent of elements in each repeat class. DNMT1-maintained and DNMT3a/3b-maintained regions included for reference. Question mark (?) indicates uncertainty in classification.

Since correlation in DNMT1-maintained regions was weakest between lower and upper dose groups, we ranked repeat classes in descending order by Pearson's r in lower versus control comparison. For LvC, Pearson's r at repeat classes ranged from 0.85 to 0.99 (adj $p < 1 \times 10^{-13}$). For UvC, Pearson's r ranged from 0.86 to 0.99 (adj $p < 1 \times 10^{-13}$). For UvL, Pearson's r ranged from 0.84 to 0.99 (adj $p < 1 \times 10^{-13}$). Median methylation change for repeat classes ranged from and increase of 0.7 percent to 2.1 percent in LvC, a decrease of 0.6 percent to an increase of 0.3 percent in UvC, and a decrease of 0.8 percent to 2.1 percent in LvC (Fig. 3.3B). This suggests that repeat classes experienced different propensities toward hypermethylation methylation at lower dose. The majority of individual members for each repeat class behaved consistently as indicated by low IQRs (Fig. 3.3C).

Among the mouse retrotransposon classes, short interspersed elements (SINEs) showed the highest correlations between dose groups (range = 0.95-0.96, N = 911,561 elements, $p < 1 \times 10^{-13}$), followed by LTRs (range = 0.91–0.93, N = 618,104, $p < 1 \times 10^{-13}$) and LINEs (range = 0.88–0.91, N = 641,607 elements, $p < 1 \times 10^{-13}$; Fig. 3.3A). Our previous work indicated that certain retrotransposon families and subtypes had higher propensity to be methylated by DNMT1 (Li et al., 2015), we therefore examined correlation and methylation change among families of the retrotransposon classes. Pearson's r for SINEs was at least 0.94, with the highest observed at tRNA-derived SINEs (range = 0.98–0.99, $p < 1 \times 10^{-13}$; Fig. 3.4A). For each comparison, median methylation change was similar across SINE families and IQRs were low (Fig. 3.4B–C). LTR retrotransposons showed a broader range of correlation coefficients (range = 0.90 to 0.97, $p < 1 \times 10^{-13}$), with the lowest of 0.90 observed at LTR and ERVL-MaLR in LvC comparison (Fig. 3.4D). Gypsy, ERVL-MaLR, ERVK, ERVL, and ERV1 had median methylation changes greater than 1.7 percent in LvC and less than -1.6 in UvL and IQRs less than 5 percent, demonstrating a relatively high propensity toward hypermethylation in lower dose group (Fig. 3.4 E–F). LINE families had the broadest range of correlation coefficients 0.79 to 0.98 ($p < 1 \times 10^{-13}$; Fig. 3.4G). L1, the most prominent LINE family in the mouse genome, had the greatest median change in methylation in LvC (+2.03%) and in UvL (-1.88%; Fig. 3.4H). LINE RTE-BovBs were resistant changes in methylation and had low IQRs in all comparisons (Fig. 3.4H–I).

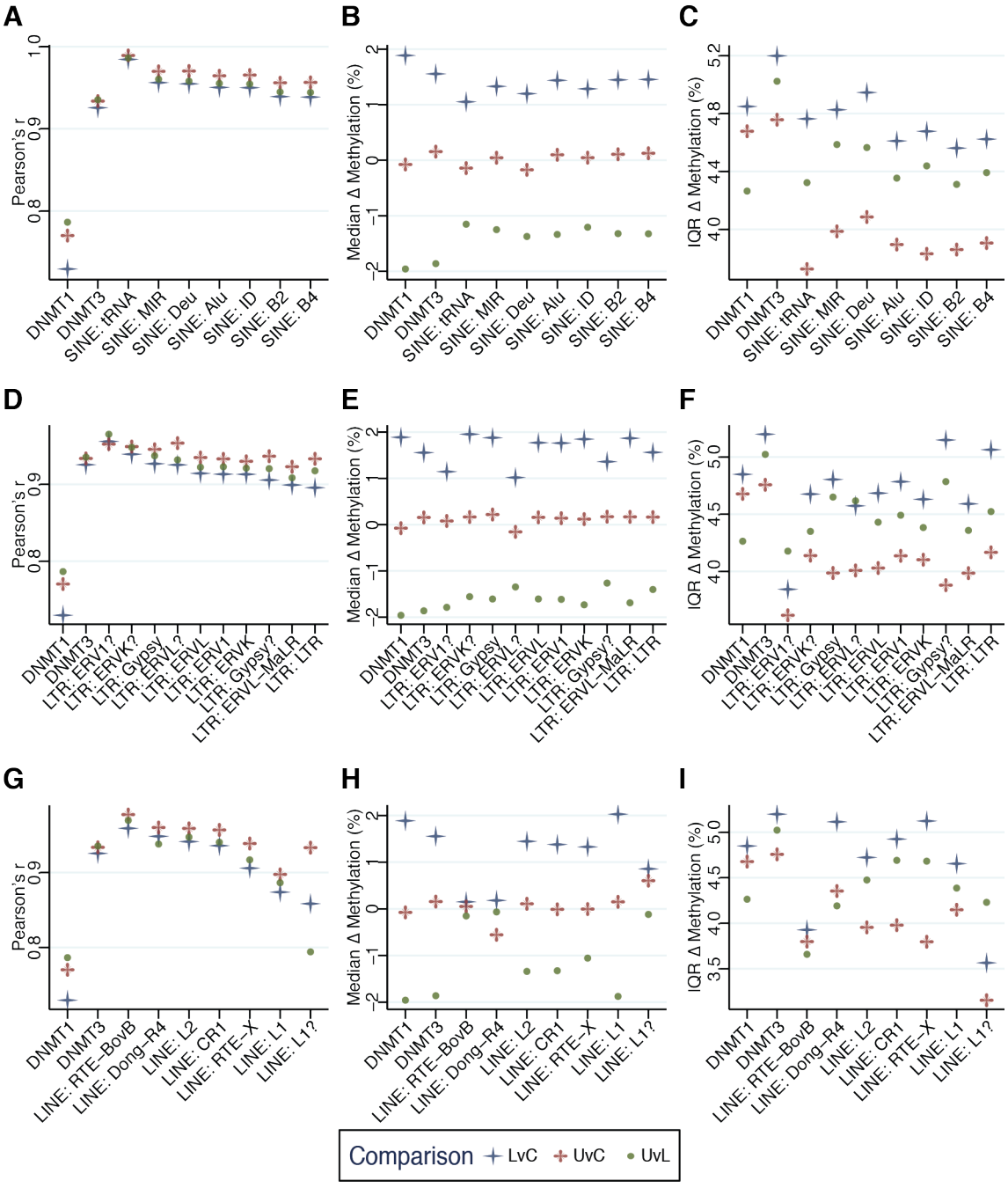
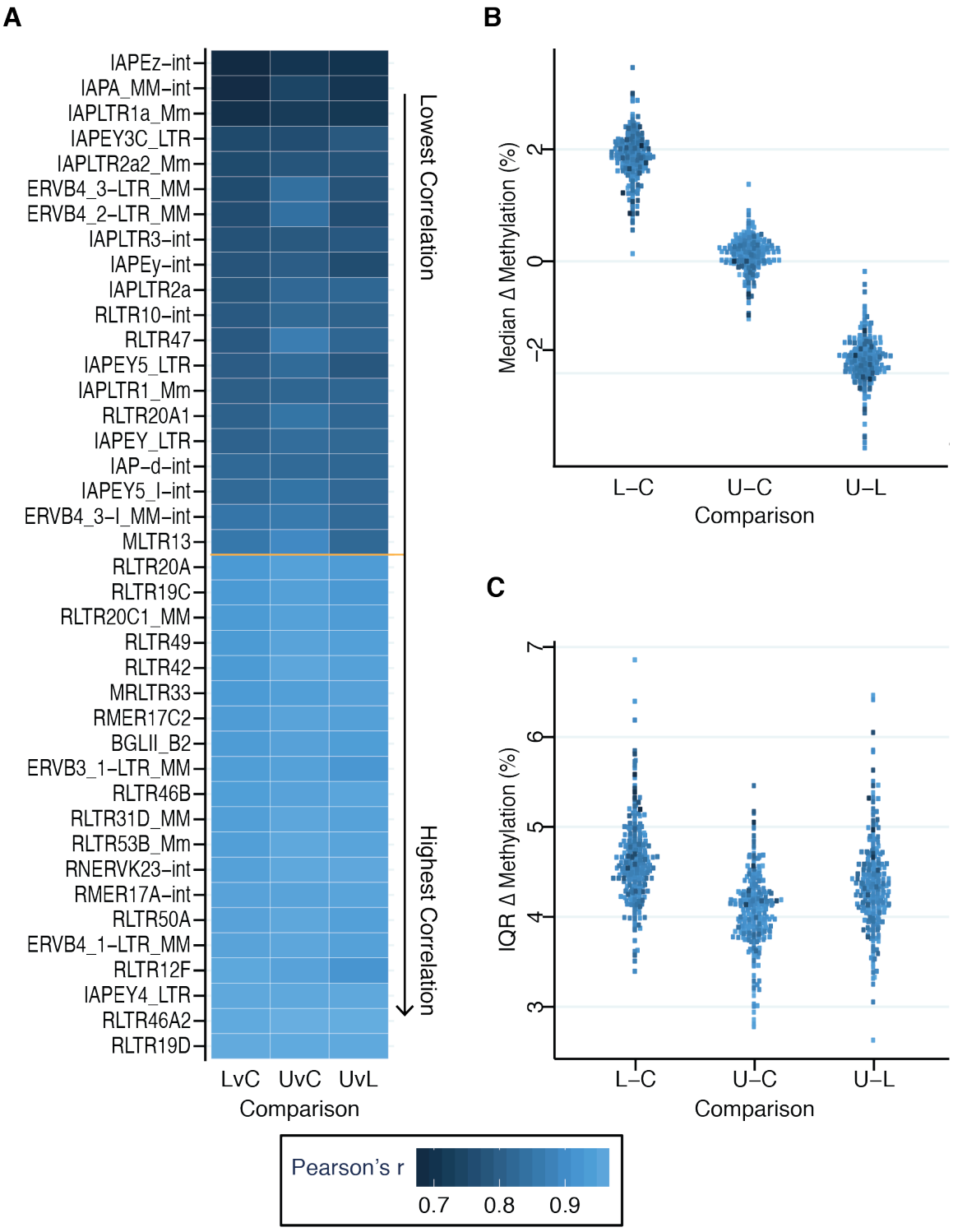


Figure 3.4 Characterization of region-level differences in methylation for mouse retrotransposon families in BPA-exposed and control fetal livers. (A–I) Correlation coefficients **(A, D, G)**, median **(B, E, H)** and interquartile range (IQR) **(C, F, I)** for change (Δ) in methylation at SINE **(A–C)**, LTR **(D–F)**, and LINE **(G–I)** families. Three comparisons are lower vs. control (LvC), upper vs. control (UvC), and upper vs. lower (UvL). IQR measures the dispersion of the middle 50 percent of elements in each retrotransposon class. DNMT1-maintained and DNMT3a/3b-maintained regions included for reference. Question mark (?) indicates uncertainty in classification.

LTR subtypes within the ERVK family have been associated with metastable epialleles (Dolinoy et al., 2007a) and display high DNMT1-dependence (Li et al., 2015; Okano et al., 1999; Tsumura et al., 2006). For ERVK LTR subtypes, Pearson's r ranged from 0.68 to 0.98. Further examination revealed 17 ERVK retrotransposon subtypes with correlation coefficients at or below 0.80 in any comparison (Fig. 3.5A). In LvC, 57.1 percent had a median change in methylation of at least 1.89, which indicates that the majority of ERVK subtypes had higher overall methylation increase than DNMT1-maintained regions (Fig. 3.5B). Many ERVK subtypes also exhibited more variation in methylation change at lower dose than DNMT1-maintained regions; IQRs were higher than 4.85 for 35.3 percent of ERVK subtypes (Fig. 3.5C). Among the LTR ERVK subtypes with relatively low correlation coefficients were IAPEz (LvC: r = 0.68, N = 5632 elements, adj $p < 1 \times 10^{-13}$) and IAPLTR1 Mm (LvC: r = 0.80, N = 1450 elements, adj $p < 1 \times 10^{-13}$; Fig. 3.5A). Closer inspection of IAPEz elements indicates a median methylation increase of 2.1 percent and an IQR of 5.2 percent in LvC (Supplemental Fig. S3.6). Interestingly, closer inspection of IAPLTR1 Mm elements reveals a median methylation increase of only 0.85 percent, which was the sixth lowest increase among ERVK family elements, and an IQR of 5.3 percent in LvC, which was the eighteenth highest among the 276 ERVK element subtypes examined (Supplemental Fig. S3.7). The majority of IAPLTR1 Mm elements had decreased methylation in upper compared to lower dose; the median decease in methylation at upper dose was sixth among greatest among the ERVK family.

Figure 3.5 Characterization of region-level differences in methylation for ERVK family LTR retrotransposons in BPA-exposed and control fetal livers. (A) Heatmap of correlation coefficients for methylation level at ERVK LTR subtypes for lower compared to control (LvC), upper compared to control (UvC), and upper compared to lower (UvL) dose groups. Twenty subtypes with lowest correlation coefficients in LvC comparison are shown above line. Twenty subtypes with highest correlation coefficients in LvC comparison are shown below line. **(B-C)** Region-level methylation differences (Δ): each point represents the median **(B)** and interquartile range (IQR) **(C)** of methylation difference for elements in each subtype (i.e., each subtype has one point). IQR measures the dispersion of the middle 50 percent of elements for each subtype.


A
IAPEz-int
IAPA_MM-int
IAPLTR1a_Mm
IAPEY3C_LTR
IAPLTR2a2_Mm
ERVB4_3-LTR_MM
ERVB4_2-LTR_MM
IAPLTR3-int
IAPEy-int
IAPLTR2a
RLTR10-int
RLTR47
IAPEY5_LTR
IAPLTR1_Mm
RLTR20A1
IAPEY_LTR
IAP-d-int
IAPEY5_I-int
ERVB4_3-I_MM-int
MLTR13
RLTR20A
RLTR19C
RLTR20C1_MM
RLTR49
RLTR42
MRLTR33
RMER17C2
BGLII_B2
ERVB3_1-LTR_MM
RLTR46B
RLTR31D_MM
RLTR53B_Mm
RNERVK23-int
RMER17A-int
RLTR50A
ERVB4_1-LTR_MM
RLTR12F
IAPEY4_LTR
RLTR46A2
RLTR19D
Lowest Correlation
Highest Correlation
LvC
UvC
UvL
Comparison
B
Median Δ Methylation (%)
2
0
−2
L−C
U−C
U−L
Comparison
C
IQR Δ Methylation (%)
7
6
5
4
3
L−C
U−C
U−L
Comparison
Pearson's r
0.7
0.8
0.9

Correlation among LTR subtypes within the ERVL-MaLR family was generally homogenous (Fig. 3.6A). The lowest correlation among ERVL-MaLR elements was observed in MLT1E-int; however, this family had five representative elements and the correlation was not significant (LvC: r = 0.4, adj p = 1; UvL: r = 0.01, adj p = 1). For ERVL-MaLR subtypes with at least ten representative elements, correlation coefficients ranged from 0.77 to 0.98; the lowest was observed for MLT1G1 in UvC comparison (r = 0.77, adj p = 1.2 x 10^{-4}, N = 32 elements). In LvC, the majority of ERVL-MaLR subtypes, 52.8 percent, had a median change in methylation lower than that observed for DNMT1-maintained regions (Fig. 3.6B). ERVL-MaLR exhibited less variation in methylation change at lower dose than DNMT1-maintained regions; IQRs were higher than 4.85 for only 17.92 percent subtypes (Fig. 3.6C).

Finally, we examined L1 subtypes. Pearson's r ranged from 0.76 to 0.96. Only one subtype L1M3a had a correlation coefficient less than 0.80 (r = 0.76, adj p = 8.13 x10^{-12}, N = 64 elements; Fig. 3.7A). In LvC, 48.9 percent and 77.10 percent of L1 subtypes had higher median methylation change than DNMT1-maintained and DNMT3a/3b-maintained regions, respectively. Only 15.3 percent and 3.1 percent of L1 subtypes had IQRs higher than those for DNMT1-maintained and DNMT3a/3b-maintained regions.

Figure 3.6 Characterization of region-level differences in methylation for ERVL-MaLR family LTR retrotransposons in BPA-exposed and control fetal livers. (A) Heatmap of correlation coefficients for methylation level at ERVL-MaLR LTR subtypes for lower compared to control (LvC), upper compared to control (UvC), and upper compared to lower (UvL) dose groups. Twenty subtypes with lowest correlation coefficients in LvC comparison are shown above line. Twenty subtypes with highest correlation coefficients in LvC comparison are shown below line. **(B-C)** Region-level methylation differences (Δ): each point represents the median **(B)** and interquartile range (IQR) **(C)** of methylation difference for elements in each subtype (i.e., each subtype has one point). IQR measures the dispersion of the middle 50 percent of elements for each subtype.

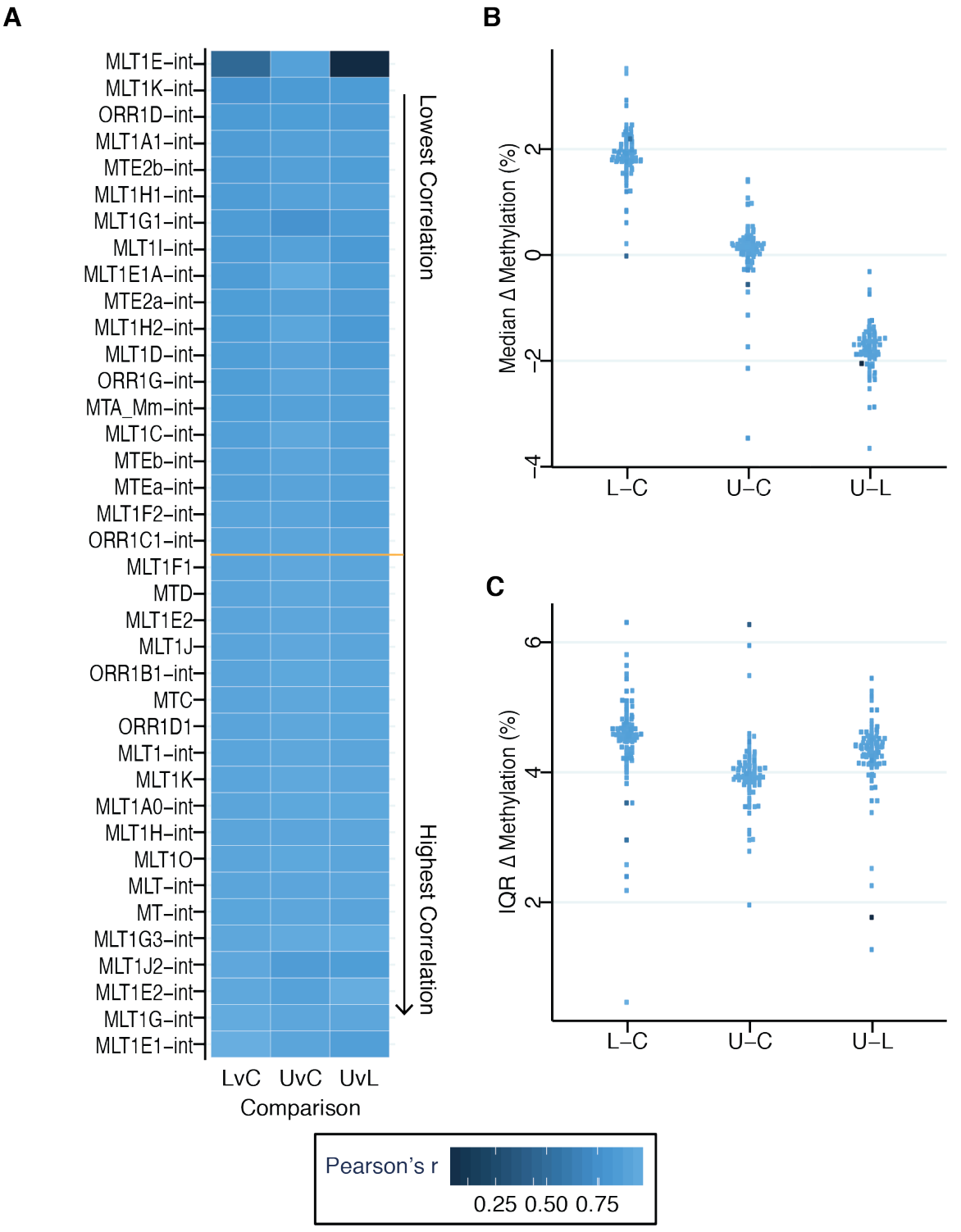
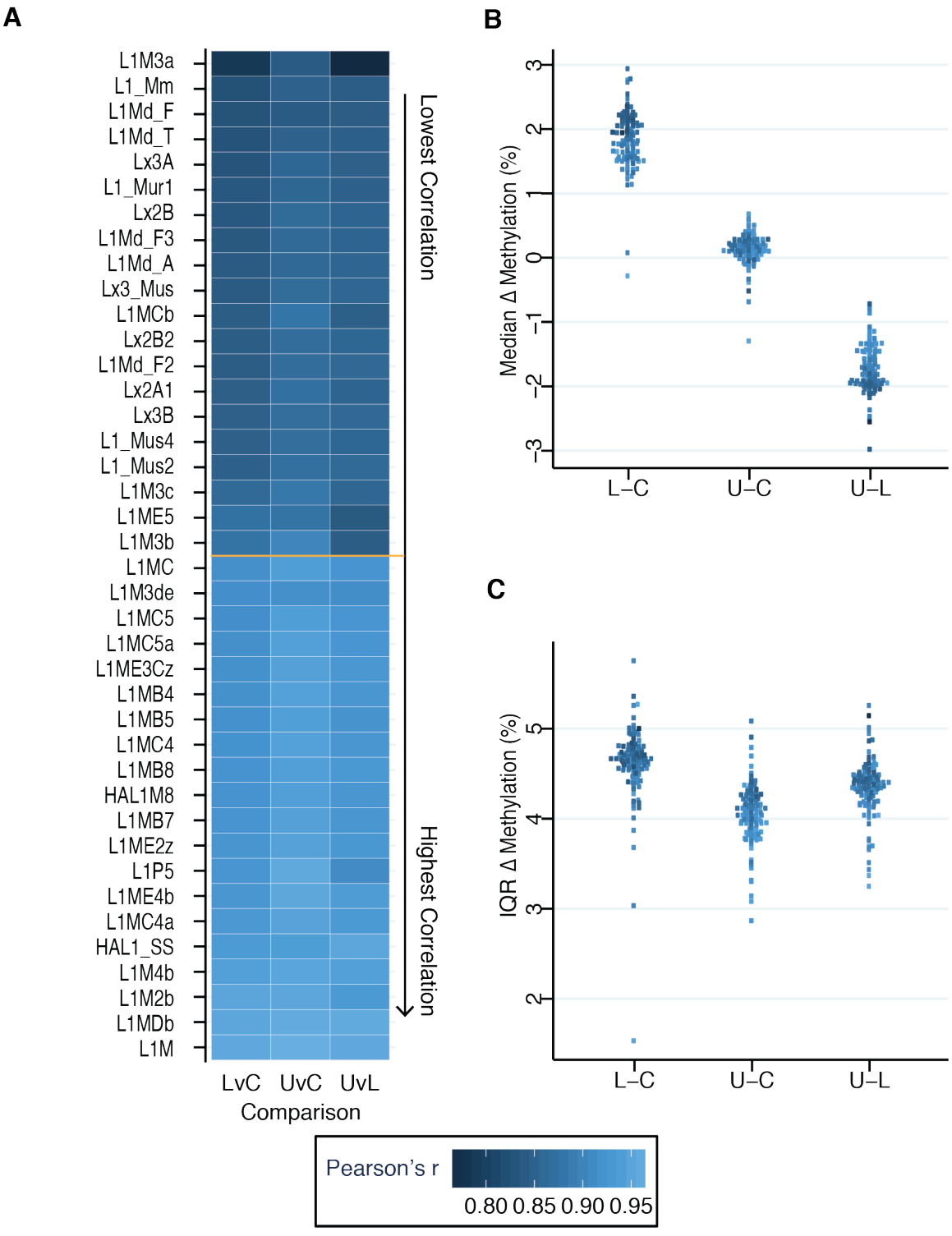

A
MLT1E-int
MLT1K-int
ORR1D-int
MLT1A1-int
MTE2b-int
MLT1H1-int
MLT1G1-int
MLT1I-int
MLT1E1A-int
MTE2a-int
MLT1H2-int
MLT1D-int
ORR1G-int
MTA_Mm-int
MLT1C-int
MTEb-int
MTEa-int
MLT1F2-int
ORR1C1-int
MLT1F1
MTD
MLT1E2
MLT1J
ORR1B1-int
MTC
ORR1D1
MLT1-int
MLT1K
MLT1A0-int
MLT1H-int
MLT1O
MLT-int
MT-int
MLT1G3-int
MLT1J2-int
MLT1E2-int
MLT1G-int
MLT1E1-int
Lowest Correlation
Highest Correlation
LvC
UvC
UvL
Comparison
B
Median Δ Methylation (%)
2
0
−2
−4
L−C
U−C
U−L
C
IQR Δ Methylation (%)
6
4
2
L−C
U−C
U−L
Pearson's r
0.25 0.50 0.75

Figure 3.7 Characterization of region-level differences in methylation for LINE L1 retrotransposons in BPA-exposed and control fetal livers. (A) Heatmap of correlation coefficients for methylation level at LINE L1 subtypes for lower compared to control (LvC), upper compared to control (UvC), and upper compared to lower (UvL) dose groups. Twenty subtypes with lowest correlation coefficients in LvC comparison are shown above line. Twenty subtypes with highest correlation coefficients in LvC comparison are shown below line. **(B-C)** Region-level methylation differences (Δ): each point represents the median **(B)** and interquartile range (IQR) **(C)** of methylation difference for elements in each subtype (i.e., each subtype has one point). IQR measures the dispersion of the middle 50 percent of elements for each subtype.


A
L1M3a
L1_Mm
L1Md_F
L1Md_T
Lx3A
L1_Mur1
Lx2B
L1Md_F3
L1Md_A
Lx3_Mus
L1MCb
Lx2B2
L1Md_F2
Lx2A1
Lx3B
L1_Mus4
L1_Mus2
L1M3c
L1ME5
L1M3b
L1MC
L1M3de
L1MC5
L1MC5a
L1ME3Cz
L1MB4
L1MB5
L1MC4
L1MB8
HAL1M8
L1MB7
L1ME2z
L1P5
L1ME4b
L1MC4a
HAL1_SS
L1M4b
L1M2b
L1MDb
L1M
Lowest Correlation
Highest Correlation
LvC
UvC
UvL
Comparison
B
Median Δ Methylation (%)
3
2
1
0
−1
−2
−3
L−C
U−C
U−L
C
IQR Δ Methylation (%)
5
4
3
2
L−C
U−C
U−L
Pearson's r
0.80 0.85 0.90 0.95

3. Discussion

DNA methylation characteristics at imprint and imprint-like regions increase these regions' susceptibility to BPA-induced epigenetic perturbation

Based on previous studies, it was not surprising to find altered methylation at several imprinted genes. Perinatal exposure to BPA has been shown to induce methylation changes and/or reduce parent-of-origin-dependent allelic bias in gene expression at several imprinted genes (Iqbal et al., 2015; Kang et al., 2014; Susiarjo et al., 2013). In this study, we identified altered methylation at the gDMR for *Gtl2* (*Meg3*), *H19*, *Igf2r*, *Kcnq1ot1*, *Nespas-Gnas*XL*, Peg1*, *Peg10*, *Peg13*/*Trappc9*, *Rasgrf1*, *Snrpn*, and *Zrs1*/*Commd1*(*Murr1*). Interestingly, within several known gDMRs, we observed hypomethylation and hypermethylation of CpG sites, depending on the location within the larger region. This could indicate local dysregulation of methylation that controls specific genes or transcript variants within an imprinted region.

In addition to gDMRs, we observed altered methylation at bimodal NORED regions within imprinted gene clusters. We identified altered CpG sites near seven gDMRs that had altered methylation and near one gDMR, *H13*/*Mcts2*, where we did not observe changes within the associated gDMR. This likely reflects site-specific regulation of imprinted gene clusters. For example, the *H19* promoter is considered a somatic DMR (sDMR) that does not exhibit differential methylation between sperm and oocyte, but develops parent-of-origin dependent methylation during early embryonic development (Thorvaldsen et al., 2002). For *H19*, the gDMR and sDMR are typically methylated on the paternal allele and unmethylated on the maternal allele (Thorvaldsen et al., 2002). Therefore, the increase in methylation at the *H19* promoter could be indicative of aberrant methylation gain on the maternal allele.

Altered methylation outside of known gDMRs was not limited to CpG sites within known imprinted gene clusters. Prenatal BPA exposure affected CpG sites within 17 novel bimodal NORED regions. This suggests imprinted clusters and novel bimodal NORED regions share epigenetic characteristics that influence the likelihood for an altered epigenetic state in response to prenatal BPA exposure. That an increased propensity for altered methylation occurred at exclusively NORED and exclusively bimodal regions further supports the assertion that epigenetic attributes associated with non-

restorable DNMT1-dependent methylation and bimodal methylation patterns contribute to regional susceptibility to epigenetic disturbances.

DNA methylation maintenance paradigm influences BPA-altered methylation

In addition to epigenetic characteristics observed at imprinted loci and bimodal NORED regions, the distinct methylation maintenance paradigms at some loci also increase the chance that CpG sites will be altered by prenatal BPA exposure. DNMT1-maintained and DNMT3a/3b-maintained regions were more likely to have CpG sites with altered methylation in the top 99.5th percentile or bottom 0.5th percentile than genomic regions relying on the cooperativity between the enzymes for methylation maintenance. In other words, methylation patterns maintained by both maintenance DNMT1 and *de novo* DNMT3a/3b were more resistant to perturbation.

Both loss of DNA methylation through disruption of maintenance and gain of methylation through *de novo* methylation can contribute to aberrant DNA methylation patterns in response to toxicant exposure. Overall, CpG sites within DNMT1-maintained regions appear to follow the direction of genome-wide changes, but showed more extreme hypermethylation at lower dose. This is consistent with the idea that different classes of sequences show varying degree of susceptibility to gain of methylation from *de novo* enzymatic activities of DNMT1 (Biniszkiewicz et al., 2002). Indeed, more CpG sites were within the top 99.5th percentile (hypermethylated) than the bottom 0.5th percentile (hypomethylated) in LvC comparison and more CpG sites were within the bottom 0.5th percentile than in the top 99.5th percentile for UvC and UvL comparisons.

In contrast, many CpG sites within DNMT3a/3b regions appear to favor increased methylation at upper dose compared to control or lower dose. Despite the genome-wide shift toward higher methylation at lower dose, more CpG sites were within the top 99.5th percentile than in the bottom 0.5th percentile in UvL comparison, which indicates site-specific hypermethylation in the upper exposure group. Others have identified regions with poor DNMT1-dependent maintenance that are compensated for with *de novo* DNMT3a/3b-dependent methylation (Liang et al., 2002). Therefore, it is possible that DNMT3a/3b-dependent regions experience a probabilistic methylation loss, but targeted *de novo* methylation gain

leading to high interindividual variability for methylation loss, but low interindividual variability for methylation gain. Consistently, overexpression of different DNMT enzymes has been shown to lead to hypermethylation of distinct targets (Choi et al., 2010). Overall our findings suggest that DNA methylation maintenance paradigm as well as site-specific *de novo* methylation contribute to BPA-induced changes in methylation patterns.

Repetitive elements show variation in BPA-induced methylation changes

The distinct, dose-dependent responses of preferentially DNMT1-maintained and preferentially DNMT3a/3b-maintained regions to BPA-induced methylation changes prompted us to examine the methylation changes of repetitive elements throughout the genome. We have previously demonstrated that certain types of IAP retrotransposons are resistant to methylation loss in the absence of DNMT3a/3b (Li et al., 2015). In other words, they retain relatively high methylation with only DNMT1 present. In addition, others have shown that overexpression of *Dnmt1* in mESCs increases global methylation and leads to hypermethylation of IAP elements, including at IAP elements that had become hypomethylated in DNMT1-deficient cells, but not the CpG islands at nonimprinted genes (Biniszkiewicz et al., 2002). However, others have shown that DNMT1 alone could not restore methylation at specific sequences that had been efficiently maintained by DNMT1 prior to 5-aza-2'-deoxycytidine treatment (Liang et al., 2002). In the study, Liang et al. classified several DNA fragments as highly DNMT1-dependent (CI-f fragments) or highly DNMT3a/3b-dependent (CII-d fragments). Treatment with 5-aza-2'-deoxycytidine reduced methylation levels of both fragment types in WT mESCs. After 5-aza-2'-deoxycytidine was removed, the methylation at both CI-f and CII-d fragments was increased. In *Dnmt1*$^{-/-}$ mESCs, CI-f fragments were not remethylated after 5-aza-2'-deoxycytidine removal. Furthermore, they demonstrated that in *Dnmt3a*$^{-/-}$ /*Dnmt3b*$^{-/-}$ mESCs, methylation of CII-d fragments was reduced by 5-aza-2'-deoxycytidine treatment, but restored once 5-aza-2'-deoxycytidine was removed. In other words, DNMT3a/3b alone could restore methylation at DNMT3a/3b-dependent repetitive sequences after methylation reduction from transient treatment with 5-aza-2-deoxycytididine, whereas restoration of methylation at DNMT1-maintained regions required activity from both DNMT1 and DNMT3a/3b (Liang et al., 2002). We proposed that the maintenance and *de novo* methylation activity from the combined action of three enzymes at a given repeat

locus could lead to varied epigenetic responses among repetitive elements in response to environmental exposures. We therefore, analyzed the similarities in region-level methylation changes between dose groups for repeats at different levels of hierarchical classification from the broad category of repeat class through specific subtypes of retrotransposon families.

Among retrotransposon classes, SINE elements showed the highest correlations in region-level methylation between exposure groups (Fig. 3.3A). SINE families were relatively homogenous in terms of overall methylation changes observed, whereas LINE and LTR displayed greater heterogeneity among families (Fig. 3.4). LINE L1 family and several LTR families (Gypsy, ERVL-MaLR, ERVK, ERVL, and ERV1) exhibited the largest overall changes in methylation at lower dose, indicating that these families have high propensity for *de novo* methylation at low dose BPA exposure (Fig. 3.4E, H). ERVL-MaLR LTR subtypes were relatively homogenous in correlations between dose groups (Fig. 3.6).

LINE L1 subtypes displayed a broad range (0.20) of correlations between dose groups (Fig. 3.7). The majority of LINE L1 subtypes showed methylation increases at lower dose BPA that were higher than those of DNMT3a/3b-maintained regions, but lower than those of DNMT1-maintained regions (Fig. 3.7B–C, 3.3J). A small proportion of LINE L1 subtypes showed more variation in response to lower dose BPA than either DNMT1 or DNMT3a/3b-dependent regions (Fig. 3.7C, 3.3J). Together, this suggests that in general, LINE L1 subtypes experience *de novo* methylation at low dose BPA, LINE L1 subtypes display heterogeneity between subtypes, and most LINE L1 elements behave homogenously within a subtype.

LTR ERVK subtypes, which include IAPs, displayed a very broad range of correlation coefficients (Fig. 3.5A). The majority of subtypes had methylation gains greater than those of DNMT1-maintained regions and over one-third of subtypes showed greater variation than DNMT1-maintained regions (Fig. 3.5B–C, 3.3J). Among those with the lowest correlation coefficients between exposure groups, IAPEz-int, IAPLTR1a_Mm, and IAPLTR1_Mm have been ranked as the top three LTR subtypes for DNMT1-dependent methylation maintenance; IAPLTR3-int was rank ninth and IAPEy-int was ranked eleventh (Li et al., 2015). None of the LTRs among the 20 ERVK subtypes with highest correlations between dose groups were considered to have a high degree of DNMT1-dependence (Fig. 3.5A). While the link between dependency on DNMT1 and regional vulnerability to BPA has yet to be explored, BPA has

been shown alter *Dnmt1* levels in livers of adult male mice (Ke et al., 2016) and in zebrafish (*Danio rerio*) in a dose- and tissue-specific manner (Laing et al., 2016).

In addition to displaying a proclivity for DNMT1-dependent maintenance, IAPLTR1 Mm elements have been associated with metastable epialleles (Faulk et al., 2013) that are responsive to epigenetic perturbation from environmental and nutritional stressors, including BPA (Dolinoy et al., 2006; 2007b). Intriguingly, IAPLTR1 Mm elements were more resistant to *de novo* methylation gain at lower dose BPA and tended to have a higher propensity for methylation loss at upper dose levels than other ERVK LTRs examined. Although, not evaluated on a genome-wide basis, studies using targeted methods to compare the degree of intraspecies variation at specific IAPLTR1 Mm loci have shown that individual copies of IAPLTR1 Mm demonstrate substantial intraspecies variability in methylation levels, but that the methylation averaged across all 21 IAP elements tested for each animal is highly consistent (Faulk et al., 2013). We speculate that this could reflect a probabilistic methylation gain or loss at individual copies of IAPLTR1 Mm elements within the genome of a single organism, but an overall stable level of methylation in the absence of exposure. Consistent with the idea that "global methylation of repetitive loci does not necessarily reflect locus specific deviation" (Faulk et al., 2013), this implies that shifts in genome-wide average DNA methylation among retrotransposons may not necessarily be indicative of changes at specific loci. Indeed, individual copies of IAPLTR1 Mm elements demonstrated a substantial degree of variability in response to BPA exposure in this study. Our results support that regional epigenetic context (e.g., higher-order chromatin structure) combines with genomic context and environmental influences to determine locus-specific DNA methylation. Future work should explore the interindividual variability in genome-wide and locus-specific repetitive element responses to prenatal BPA exposure. Furthermore, studies interested in identifying loci with environment-influenced changes in DNA methylation should include examination of DNMT levels to determine their contribution to intraspecies variability.

4. Conclusions

In summary, our work has demonstrated that epigenetic properties characteristic of imprinted gDMRs and certain transposable element families increase the likelihood that a genomic region will be susceptible to epigenetic perturbation from exposure to environmental compounds. DNMT1-dependent,

non-restorable methylation loss (i.e., NORED) and bimodal methylation patterns are distinct epigenetic characteristics that occur at, but are not exclusive to, imprinted germ line DMRs. CpG sites within genomic regions that displayed either characteristic were more likely to have altered CpG site methylation compared to CpG sites genome-wide. Genomic regions displaying both characteristics, including several imprinted gDMRs, were the most likely to have altered CpG methylation. DNA methylation maintenance paradigms influenced region-level methylation changes between dose groups. DNMT1-maintained regions and retrotransposon types with high DNMT1 dependence (e.g., IAPLTR1 Mm) displayed relatively low correlation in region-level methylation levels following *in utero* exposure to physiologically relevant doses of BPA. Overall, our results imply a role for DNMT1-mediated mechanisms driving BPA-induced methylation changes at imprinted genes and certain retrotransposons. Species-specific differences in imprinted regions and repetitive elements imply potential for mechanistic concordance rather than, or in addition to, phenotypic concordance from environmental and nutritional exposures.

5. Materials and Methods

Data set is unpublished data (see Chapter 4 for library construction and data processing details). DNMT1-maintained and DNMT3a/3b-maintained regions obtained from (Li et al., 2015). NORED, bimodal, and bimodal NORED regions are from Chapter 2. Repeat classification obtained by downloading RepeatMasker (Smit et al., 2013) track for mouse mm10 genome from UCSC Table browser (Karolchik et al., 2004; Kent et al., 2002).

For each pairwise, dose-group comparison, altered CpG sites were defined as those within the top 99.5th percentile (hypermethylation) or bottom 0.5th percentile of the t-statistic distribution (i.e., 1% of CpG sites were considered altered genome-wide). ORs were calculated for CpG sites within each region having altered methylation compared to autosomal CpG sites for bimodal, NORED, and bimodal NORED regions or compared to all CpG sites for DNMT1-maintained and DNMT3a/3b-maintained regions. P-values and confidence intervals for ORs were calculated with two-sided Fisher's exact test using 'exact2x2' R package (Fay, 2010) and corrected for multiple comparisons (15 comparisons total for our other regions of interest) to control family-wise error rate below 0.05

To examine the correlation and change in methylation between dose groups, we averaged the smoothed CpG site methylation estimates across all biological replicates for control (n = 3), lower (n = 4), and upper (n = 4) exposure levels. For region-level analyses we used the 'perRegion' option of the getMeth() function from the 'bsseq' R package to obtain the average per-region methylation estimate for each replicate prior to calculating the dose group average (Hansen et al., 2012). Pearson's r was calculated and p-values were adjusted for 563 tests with Bonferroni correction.

6. List of Supplemental Materials

Supplemental Figures S3.1 to S3.7 (**Appendix II**)

Figure S3.1 Methylome characterization: bimodal regions
Figure S3.2 Methylome characterization: bimodal NORED regions
Figure S3.3 Methylome characterization: NORED regions
Figure S3.4 Methylome characterization: DNMT1-maintained regions
Figure S3.5 Methylome characterization: DNMT3a/3b-maintained regions
Figure S3.6 Methylome characterization: IAPEz
Figure S3.7 Methylome characterization: IAPLTR1 Mm

Chapter 4

DNA methylation-mediated pathway perturbation provides broad insight into Bisphenol A-induced metabolic abnormalities in male mice from prenatal exposure

(This work was completed as collaboration between the Wang Lab (Johns Hopkins University) and Bartolomei Lab (University of Pennsylvania). Bartolomei Lab conducted animal exposures, RNA-sequencing, and RNA-sequencing data analyses. SNM was responsible for DNA methylation experiments and associated analyses, except for Ingenuity Pathway Analyses. SNM wrote Results and Methods Sections pertaining to DNA methylation and contributed to Abstract, Introduction, and Discussion.)

Abstract

Increasing evidence has demonstrated that environmental exposure induced-fetal reprogramming events are linked to metabolic disorders; the precise mechanisms, however, are unclear. The goal of the study is to elucidate molecular mechanisms underlying metabolic abnormalities in male mice exposed *in utero* to 10 μg/kg/day and 10 mg/kg/day of bisphenol A (BPA). We used whole genome bisulfite sequencing and RNA sequencing to determine if exposure altered DNA methylation and/or transcription of metabolic genes. Our combined -omics approach revealed that BPA exposure produced dose-dependent alterations in the methylome and transcriptome of mouse fetal liver. BPA exposure perturbed DNA methylation and/or transcription in pathways related to metabolic homeostasis, indicating that both DNA methylation-mediated and DNA methylation-independent mechanisms may contribute to prenatal programming of adult metabolic phenotypes. Future work focusing on these pathways may provide critical insight into risk assessment and management of metabolic diseases with underlying environmental causes.

1. Introduction

Exposure to various environmental factors during pregnancy is linked to compromised fetal metabolic health, a paradigm known as the developmental origin of health and disease (DOHaD). Altered nutrients, environmental or occupational exposure, and maternal factors can significantly impact the long-term health outcome of the fetus and increase susceptibility to adult diseases including obesity, diabetes, and cardiovascular abnormalities (Barouki et al., 2012). More recently, studies further suggested that altered *in utero* programming due to exposure to environmental factors also influences health of the subsequent generations (Carone et al., 2010; Radford et al., 2014; Skinner, 2014; Xin et al., 2015).

Elucidating molecular mechanisms of early environment-induced developmental reprogramming has been one of most important goals for DOHaD researchers. Research that ultimately identifies and elucidates susceptible developmental genes, as well as the associated pathways, provides not only critical insights into the mechanisms of DOHaD, but also potentially facilitates future prevention and intervention strategies. Recent studies have reported diverse mechanisms or biomarkers that predict the health outcome of a fetus, including altered hormone levels leading to the disappearance of sexually dimorphic endpoints (Vigé et al., 2008), accelerated telomere shortening in response to stress (Entringer et al., 2011), and aberrant epigenetic modifications in response to various environmental insults (Ho et al., 2017). The latter mechanism has been a focus of the DOHaD field as increasing studies have reported the ability of environmental factors to alter DNA methylation and histone modifications at development- and disease-relevant genes in animal models and humans. Additionally, global and genome-wide analyses have started to elucidate developmental gene programming events susceptible to environmental insults, including endocrine-disrupting chemical (EDC) exposure and maternal undernutrition (Iqbal et al., 2015; Radford et al., 2014; van Esterik et al., 2014). Correlating these gene-specific and genome-wide changes to environmental exposures is an important approach for dissecting DOHaD mechanisms, and more importantly, investigating these associations in model systems with characterized developmental phenotypes is key to extracting the most useful information.

Among the relevant environmental factors, there has been enormous interest in studying epigenetic perturbations following exposure to the common EDC bisphenol A (BPA). One of the most produced chemicals worldwide, BPA is used in plastics, food and beverage packaging, and various other consumer products. Ubiquitous contamination of BPA in the environment, combined with its links to adverse health outcomes (Saal and Hughes, 2005), makes this chemical an intriguing and important model EDC to study. More importantly, as a representative environmental estrogen, studies on BPA may provide further insights into the epigenetic mechanisms linked to developmental effects induced by this class of common chemicals (McLachlan, 2016). Previous studies have shown that BPA exposure induces epigenetic misregulation at genes associated with human disease (Bromer et al., 2010; Susiarjo et al., 2013; Wong et al., 2015). More recently, Iqbal et al., (2015) conducted a genome wide DNA methylation using methylated CpG island recovery assay (MIRA) in JF1 X OG2 mice exposed to 0.2 mg/kg/day BPA and

other EDCs. The study did not detect significant DNA methylation changes on *de novo* DNA methylation at CpG islands, promoters, imprinted differentially methylated regions (DMRs), intracisternal A particle (IAP) retrotransposons, and along the Y chromosome, although the exposure paradigm had not been shown to induce significant gene expression changes at DNA methylation-regulated developmental genes including at imprinted genes (Kang et al., 2011). In a separate study, Van Esterik et al., (2015) performed the digital restriction enzyme analysis of methylation (DREAM) in adult female C57BL/6 X FVB mouse liver exposed *in utero* to 3000 µg/kg bw/day BPA. The study found few differentially methylated genes that are potentially linked to adult metabolic phenotypes. Given that BPA exposure has been significantly linked to metabolic phenotypes (Alonso-Magdalena et al., 2015) and that epigenetic misregulations are associated with its exposure (Ferreira et al., 2014; Singh and Li, 2012; Xin et al., 2015), most genome-wide DNA methylation studies to date have not provided satisfying mechanistic insights into the link between exposure and phenotypes.

Previously, our laboratory demonstrated that *in utero* exposure of C57BL/6 mice to 10 mg/kg bw/day and 10 µg/kg bw/day BPA altered DNA methylation patterns at imprinted control regions (ICRs) and caused aberrant expression of imprinted genes (Susiarjo et al., 2013). We subsequently showed that gestational exposure in F0 female mice produced glucose intolerance and abnormal insulin function in F1 and F2 adult male offspring (Susiarjo et al., 2015). Using a similar exposure paradigm, in this study we conducted whole-genome bisulfite sequencing (WGBS) and RNA sequencing (RNA-seq) on livers from embryonic day (E) 18.5 male mice, to gain a broad perspective on potential molecular mechanisms underlying the prenatal programming of adult metabolic phenotypes. By examining liver tissue from control and BPA exposed late gestation F1 mouse fetuses, we have identified relevant genome-wide methylation and gene expression changes from prenatal exposure that occur prior to birth in a tissue crucial to adult metabolic homeostasis.

We have used this model system in the current global and genome-wide studies to elucidate how early developmental exposure to the representative EDC BPA may predispose an individual to metabolic disease later in life. BPA exposure was linked to altered expression of developmentally important genes and epigenetic changes in late gestation (E18.5) F1 mouse fetuses. Changes to gene expression do not

necessarily require methylation changes. As such, another goal of the study was to understand how environmental perturbation influenced gene expression through DNA methylation-mediated and DNA methylation-independent pathways. Our results demonstrated that BPA exposure was linked to dose-dependent alterations of the methylome and transcriptome and that our multi-omics approach suggested that alterations in lipid metabolism explained partly the perturbed metabolic homeostasis phenotype in this mouse model.

2. Results

BPA exposure results in dose-specific alteration of genome-wide mRNA expression

Distinct biochemical profiles of BPA-exposed mouse fetuses, especially in the upper dose exposure group, suggested that *in utero* exposure to environmental chemicals significantly altered metabolic programming as early as late gestation (Susiarjo et al., 2017). Previously, we demonstrated that early life BPA exposure was linked to metabolic abnormalities in the male mice during adulthood (Susiarjo et al., 2015). We showed that the metabolic phenotype was partially associated with increased expression of the imprinted insulin-like growth factor 2 (*Igf2*) gene (Susiarjo et al., 2015). Our metabolomics data demonstrated that early metabolic effects of BPA exposure involved alteration of various important biochemical pathways, further suggesting that the mechanisms leading to adulthood disease were far more complex than single gene-based pathways (Susiarjo et al., 2017). We subsequently performed RNA-seq to identify genes that were relevant to BPA exposure-associated metabolic phenotypes by searching for differentially expressed genes (DEGs) between BPA-exposed and control mice. The RNA-seq experiment was conducted on individual E18.5 fetal livers (n=3 per exposure group), representing a subset of samples that were analyzed in the metabolomics described in (Susiarjo et al., 2017). The number of reads per sample in our RNA-seq study ranged from 35-72 X 10^6 with approximately 90% of the reads aligned to the reference sequences (Supplemental Table S4.1). The mean of insert length in our fetal liver samples ranged from 127 to 139 bp, and the aligned sequences were distributed similarly across the different samples, with approximately 53-64% and 25-33% to coding regions and untranslated regions (UTR), respectively (Supplemental Table S4.1). Hierarchical clustering of biological replicates, designated as control 1–3 (C1–C3), lower 1–3 (L1–L3), and upper dose 1–3 (U1–3), showed that DEGs in C1–C3 were distinctively

grouped together and were more similar to each other relative to DEGs in L1–L3 or U1–U3 (Fig. 4.1A). This clustering, however, indicated that although L1 and L3 were similar to each other, L2 showed more similarity to the upper dose BPA biological replicates (Fig. 4.1A), indicating interindividual variation among the lower dose exposure group.

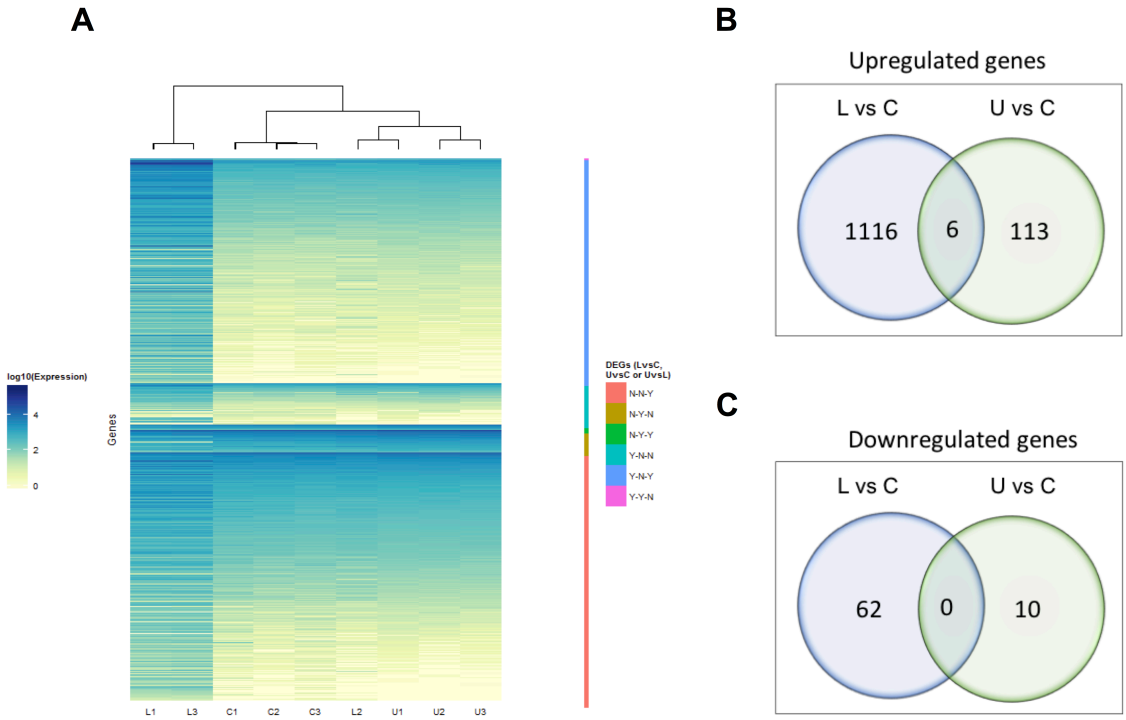


Figure 4.1 Characterization of BPA-induced mRNA expression changes. (A–C) Lower vs. control (L vs C), upper vs. control (U vs C), and upper vs. lower (U vs L) dose group comparisons. **(A)** Hierarchical clustering of methylation in biological replicates at DEGs: Y-N-N indicates that region was identified as differentially methylated in L vs C, but not in U vs C or U vs L; N-N-Y indicates that region was not considered differentially methylated in L vs C or U vs C, but identified as a DEG in U vs L; etc. **(B–C)** Overlap of genome-wide upregulated DEGs **(B)** and downregulated DEGs **(C)** among comparisons.

In contrast to metabolomics data that showed mild metabolic changes induced by lower dose BPA exposure (Susiarjo et al., 2017), our RNA-seq study demonstrated clearly that lower dose exposure produced significantly more differentially expressed genes (DEGs) than upper dose: 1184 genes were differentially expressed relative to controls, representing 1122 and 62 upregulated and downregulated genes, respectively (Fig. 4.1B–C and Supplemental Table S4.2). In the upper dose exposure group, we found 129 DEGs relative to control fetal liver (Fig. 4.1B–C and Supplemental Table S4.2) in which 119 and 10 exhibited increased and decreased expression, respectively (Fig. 4.1C). The 1184 and 129 DEGs

observed in the lower dose and upper dose BPA, respectively, were mostly mutually exclusive, as we observed only 6 overlapping DEGs (2610015P09Rik, *Arid4b*, BC005561, *Cfap36, Serf1* and *Socs2*). Intrigued by the significantly larger DEGs in the lower dose, we also compared genome-wide gene expression of fetal liver samples in the upper dose exposure group directly to the lower dose and found 2123 DEGs (Supplemental Table S4.2), representing 161 and 1962 upregulated and downregulated genes in the upper BPA dose samples, respectively. The data suggest that each BPA dose exposure produced very distinct effects on mouse fetal liver transcriptome.

For insight into relevant biological pathways affected in our BPA mouse model, we utilized the Ingenuity Pathway Analysis (IPA). IPA revealed that the DEGs were involved in various canonical pathways and toxicological functions (Supplemental Tables S4.3–S4.5). Comparison of the top 10 canonical pathways, biological, and toxicological functions (Supplemental Table S4.6) in all pairwise comparison groups (i.e., control vs. upper, control vs. lower, and upper vs. lower) showed that distinct biological pathways were perturbed in each exposure group. We observed only two overlapping pathways (Microphage migration inhibitory factor or MIF regulation of innate immunity and *Wnt* signaling pathway) in both the control vs. lower dose and upper vs. lower dose comparisons, with common genes including *Jun*, *Pla2g1b*, *Pla2g4f*, *Nos2*, *Sox11*, *Sox17*, *Rarg*, *Wnt2*, *Wnt5a*, and *Sfrp2* genes.

Linking the RNA-seq data to results from our metabolomics (Susiarjo et al., 2017), we searched for biological pathways and genes relevant to lipid, amino acid, peptide, or glucose metabolism that were altered in BPA-exposed fetal liver. Some potentially relevant pathways that differed between control and upper dose exposure, in which we saw the more significant changes of the metabolome, included cholesterol biosynthesis pathway, fatty acid metabolism, and the FXR activation known to be the limiting rate step in bile acid production (Supplemental Table S4.4). These analyses suggest that altered metabolomic profiles in BPA-exposed mice were linked to misregulation of genes involved in lipid metabolism.

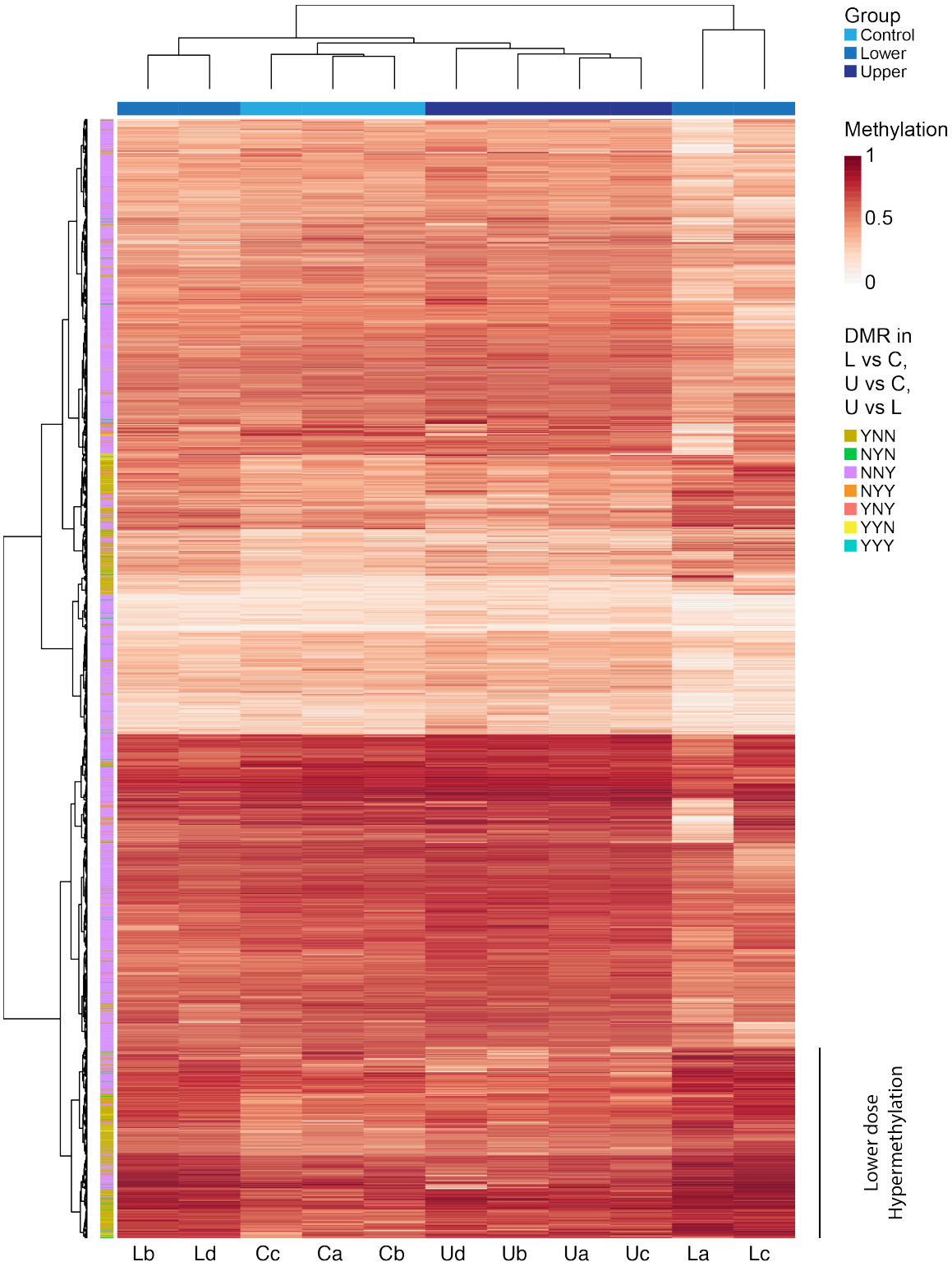
Global epigenomic profiling of in utero BPA-exposed fetal livers reveals mechanistic insights

Epigenetic marks, including DNA methylation, regulate stage- and tissue-specific gene expression and can be mitotically inherited. Altered epigenetic marks due to early life environmental exposures is a potential molecular mechanism underlying exposure-related phenotypic consequences later in life. To determine how altered DNA methylation contributes to mechanisms of early life BPA exposure-induced metabolic abnormalities (Susiarjo et al., 2015; 2017), we conducted WGBS in the liver of control (n = 3; designated Ca, Cb, and Cc), lower dose (n = 4; designated La, Lb, Lc, and Ld), and upper dose (n = 4; designated Ua, Ub, Uc, and Ud) BPA-exposed E18.5 male fetuses to generate base-resolution methylomes (see Materials and Methods). Methylomes were generated for a subset of samples that were analyzed in the metabolomics described in (Susiarjo et al., 2017). We generated 2.47 billion paired-end reads and retained 1.31 billion deduplicated, unique best hit paired-end alignments for an average of 120 million paired-end reads per sample. Median methylation level of CpG sites ranged from 66.6 percent to 70.2 percent for control fetal livers, 68.4 percent to 75.5 percent for lower dose, and 67.7 percent to 69.9 percent for upper dose (Supplemental Fig. S4.1A). Median coverage for CpG sites on autosomal and sex chromosomes ranged from 7X to 9X for control, 3X to 7X for lower dose BPA, and 5X to 15X for upper dose BPA tissues (Supplemental Fig. S4.1B–L).

To identify altered methylation patterns among BPA-exposed tissues, we used the statistical model described in (Hansen et al., 2012) that performs well on low-coverage WGBS data (Hansen et al., 2014; Harten, 2015; Ziller et al., 2015). We used the method to perform local-likelihood smoothing to increase the accuracy of methylation estimates, calculate modified t-statistics, and combine single-CpG sites into differentially methylated regions (DMRs; see Materials and Methods). To control the FDR of DMRs, we used empirical null modeling (Efron, 2007). Notably, local false discovery rates (lfdrs) were more stringent than q-values (Supplemental Fig. S4.2J–O). Our WGBS analysis showed that lower dose BPA exposure produced more DMRs (892) in fetal liver relative to upper dose (152) when compared to controls. Furthermore, we detected substantially more (2603) DMRs when upper dose was compared to lower dose BPA relative to other pairwise comparisons (Supplemental Table S4.7).

To compare individual methylation differences among fetal livers from all exposure groups, we performed hierarchical clustering of methylation values for individual replicates from each exposure group at merged DMRs (i.e., DMRs identified in all pairwise comparisons with overlapping DMRs merged into a single DMR; see Materials and Methods), which revealed four distinct groups (Fig. 4.2). One group consisted of all control replicates. A second group had all upper dose replicates. Interestingly, the lower dose replicates formed two distinct clusters, with each cluster consisting of two replicates (Fig. 4.2). Control replicates were the most similar to each other, followed by upper dose replicates. Upper dose replicates were more similar to control than to any lower dose replicates. Intriguingly, the lower dose replicates Lb and Ld were more similar to the control and upper samples than to La and Lc samples. La and Lc were distant from all other samples, which suggests a bimodal response pattern at lower BPA exposure.

Figure 4.2 Hierarchical clustering of methylation in biological replicates at merged DMRs indicates dose-dependent responses to prenatal BPA and heterogeneity among lower dose replicates. Lower vs. control (L vs C), upper vs. control (U vs C), and upper vs. lower (U vs L) dose group comparisons. YNN indicates that region was identified as differentially methylated in L vs C, but not in U vs C or U vs L; NNY indicates that region was not considered differentially methylated in L vs C or U vs C, but identified as a DMR in U vs L; etc. Lower dose BPA exposure produced 892 DMRs and upper dose exposure produced 152 DMRs when compared to controls. Upper dose versus lower dose produced 2603 DMRs. Control dose replicates form a single cluster. Upper dose replicates form a single cluster. Lower dose replicates display heterogeneity, forming two groups with two replicates in each group. The cluster of two lower dose replicates, La and Lc (right), is distinct from all other samples.


Group
Control
Lower
Upper
Methylation
1
0.5
0
DMR in
L vs C,
U vs C,
U vs L
YNN
NYN
NNY
NYY
YNY
YYN
YYY
Lower dose
Hypermethylation
Lb
Ld
Cc
Ca
Cb
Ud
Ub
Ua
Uc
La
Lc

Although our clustering of samples by DMRs from the pairwise comparisons indicated substantial methylation differences among tissues from the lower dose exposure mice, closer examination revealed that two samples within each lower dose cluster have similar changes of methylation patterns at a subset of DMRs. For example, all samples from the lower dose group shared similar dark lines (i.e., hypermethylated regions; Fig. 4.2), and this subset of DMRs appear to form three dose-dependent groups, reinforcing our confidence with the quality of our epigenomic profiling. Furthermore, each pairwise comparison produced relatively unique genome-wide DMRs, as evidenced by the relatively few shared DMRs (Fig. 4.3A). Similarly, the subset of DMRs classified as "promoter only" displayed minimal overlap between the three pairwise comparisons (Fig 4.3B). Consistent with our gene expression data, our DNA methylation data demonstrate that lower and upper dose BPA exposure produced distinct responses and that lower dose replicates displayed heterogeneity, suggestive of interindividual variability.

Given potentially different roles of intergenic, promoter, and gene body methylation patterns in maintaining genome stability and regulating gene expression (Li et al., 2015; Lorincz et al., 2004; Seisenberger et al., 2010; Stein et al., 1982; Wagner et al., 2008; Walsh et al., 1998; Wu et al., 2010; Yang et al., 2014), we examined the distribution of DMRs at intergenic and genic regions. While the different doses produced relatively unique DMRs, the DMRs were similarly distributed based on their genomic region classifications of intergenic, promoter only, promoter and gene body, and gene body only (Fig. 4.3C). For all comparisons, we observed that the majority of the DMRs were hypermethylated (Fig. 4.3D). Although all four genomic classifications showed more hypermethylated DMRs, we noted that intergenic regions had the highest proportion of hypomethylated DMRs in all comparisons (Fig. 4.3E).

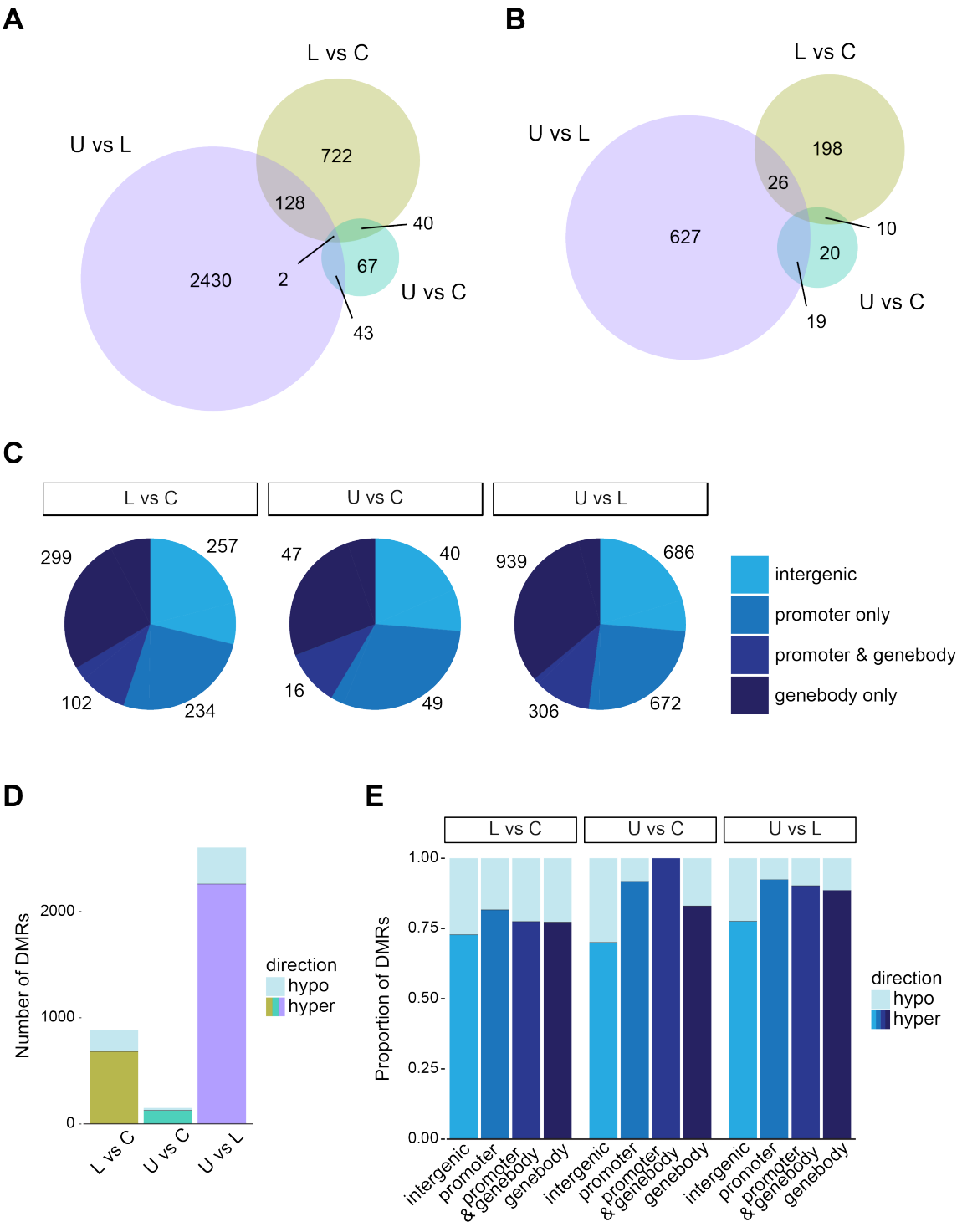


Figure 4.3 Characterization of BPA-induced methylation changes in pairwise comparisons. (A–E) Lower vs. control (L vs C), upper vs. control (U vs C), and upper vs. lower (U vs L) dose group comparisons. **(B–C)** Overlap of genome-wide DMRs **(B)** and promoter only DMRs (i.e., DMRs within 2kb upstream to 1 kb downstream a transcription start site that do not overlap the gene body of another transcript) **(C)** among comparisons. **(D)** Distribution of DMRs among genomic regions: intergenic, promoter only, promoter & gene body, and gene body only. **(E)** Number of hypo- and hypermethylated DMRs. **(F)** Proportion of hypo- and hypermethylated DMRs within each genomic feature.

Previously, studies have suggested that genomic context and CpG density influence the functional impact of CpG methylation (Blattler and Farnham, 2013; Li et al., 2015). CpG Islands (CGIs) are genomic regions with high CpG density and high GC content (Gardiner-Garden and Frommer, 1987). Cancer and tissue-specific methylation differences tend to occur in the 2 kilobase (kb) regions flanking CGIs, known as CGI shores (Materials and Methods), rather than within CGIs (Irizarry et al., 2009). In this study, DMRs from BPA-exposed tissues were significantly nearer to CGIs than expected ($p < 0.01$) and highly significantly overrepresented at CGI shores (binomial test, $p < 1 \times 10\text{-}10$), and CGI shelves (binomial test, $p < 0.01$; Materials and Methods) for all pairwise comparisons (Fig. 4.4A–B). DMRs at CGIs were slightly enriched in the lower (binomial test, $p < 0.01$, observed/expected = 2.35) and upper dose (binomial test, $p < 0.05$, observed/expected = 2.44) vs. control comparisons, but not in the upper vs. lower dose BPA (binomial test, $p = 0.1$). For CGI shores, DMRs were highly enriched when tissues from lower (observed/expected = 20.1) and upper (observed/expected = 21.9) doses were compared to controls; fewer DMRs at CGI shores were observed when the upper dose BPA group was compared to lower dose (observed/expected = 19.1). In contrast, DMRs at CGI shelves were moderately enriched when the lower (observed/expected = 3.42) and upper (observed/expected = 4.47) dose groups were compared to controls and slightly enriched in the upper vs. lower dose comparison (observed/expected = 2.46). Overall, the DMRs were similarly overrepresented at gene-associated regions (combined promoter and gene body regions; see Materials and Methods) for all pairwise comparisons (lower vs. control: binomial test, $p < 10\text{-}16$, observed/expected = 1.72; upper vs. control: binomial test, $p = 8.9 \times 10\text{-}16$, observed/expected = 1.74; upper vs. lower: binomial test, $p < 10\text{-}16$, observed/expected = 1.78; Fig. 4.4C).

The relationship between methylation and gene expression is complex, nonlinear and context dependent (Lay et al., 2015; Rishi et al., 2010; Wan et al., 2015). Furthermore, recent studies have revealed a role for DNA methylation in regulating the use of alternate promoters (Gifford et al., 2013; Maunakea et al., 2010) and in incorporating alternate exons into mRNA transcripts (Maunakea et al., 2013; Shukla et al., 2011; Yearim et al., 2015), which suggests that gene-associated DMRs could influence transcript-level rather than gene-level expression. To further characterize the genic context of BPA exposure-induced DMRs identified in this study, we asked whether the DMRs overlapped with specific categories of genic features: within 2 kb upstream of transcription start site (-2 kb TSS), within 100 bp upstream or

downstream from TSS (+/- 100 bp TSS), first exon, downstream exon, first intron, downstream intron, 5' untranslated region (UTR), 3' UTR, and coding DNA sequence (CDS) exon (see Materials and Methods). Because genic features are not mutually exclusive (e.g., a DMR within the first exon could overlap both the 3' UTR and the CDS) and DMRs can overlap different features of multiple genes or transcripts (e.g., a DMR within 2kb of the TSS for one gene could also be within the intron of an overlapping gene), we used Jaccard index and permutation tests to determine enrichment or depletion of DMRs at features along genic regions (Fig. 4.4D). DMRs for all comparisons were enriched within 2kb upstream of TSSs ($p < 0.01$), but depleted at TSSs (lower vs. control: $p < 0.01$, upper vs. control: $p < 0.05$, upper vs. lower: $p < 0.01$). DMRs were not significantly enriched at first exons for any comparison, although showed a trend toward enrichment at first exons ($p < 0.10$) in upper vs. lower DMRs. DMRs were enriched at first introns in lower vs. control ($p < 0.05$) and lower vs. upper ($p < 0.01$) DMRs. DMRs were depleted at downstream exons ($p < 0.01$), downstream introns ($p < 0.01$), and CDS exons ($p < 0.01$, $p < 0.05$, $p < 0.01$) for all comparisons. DMRs were also depleted from 5' UTRs in lower vs. control ($p < 0.01$) and 3' UTRs in upper vs. lower ($p < 0.01$). Depletion at TSSs and lack of enrichment at first exons suggests that changes in methylation are not necessarily expected to be inversely associated with gene expression. Overall, locations of gene-associated DMRs indicate potentially complex relationships between BPA-induced methylation changes and gene regulation.

DNA methylation-mediated pathway perturbation demonstrates functional differences among dose groups

To gain insight into potential functional implications of altered DNA methylation, we integrated genic DMRs with gene expression levels from RNA-seq to explore biological pathways perturbed in BPA-exposed tissues through IPA (Supplemental Table S4.8). Nine canonical pathways were enriched (-log(p-value) > 1.3) for lower vs. control DMRs and five canonical pathways were enriched (-log(p-value) > 1.3) for upper vs. control DMRs (Supplemental Tables S4.9–S4.10). Only one pathway, the GPCR-mediated Nutrient Sensing in Enteroendocrine Cells, was shared between lower vs. control and upper vs. control DMRs (Supplemental Tables S4.9–S4.10).

Consistent with distinct responses for lower and upper dose BPA exposure, we observed many (37) enriched (-log(p-value) > 1.3) canonical pathways for upper vs. lower dose DMRs; two of these

pathways (Pyrimidine Ribonucleotides *De novo* Biosynthesis and Pyrimidine Ribonucleotides Interconversion) were shared by lower vs. control DMRs and two pathways (Sphingomyelin Metabolism and Ephrin B Signaling) were shared by upper vs. control DMRs (Supplemental Table S4.11). Furthermore, upstream regulators analyses indicated activation of RICTOR (bias-corrected z-score = 2.046) in lower compared to control and inhibition of RICTOR (bias-corrected z-score = -2.954) with activation of SCAP (bias-corrected z-score = 2.738) in upper compared to lower dose BPA livers (Supplemental Tables S4.9, S4.11). Further supporting SCAP/SREBP signaling as a distinguishing factor between lower and upper dose BPA responses, gene set analysis indicated that promoter DMRs from the upper vs. lower dose comparison were enriched for genes with SREBP1 binding motifs (TCANNTGAY, q-value = 1.32 x 10^{-5}; NATCACGTGAY, q-value = 3.07 x 10^{-5}; Supplemental Table S4.12).

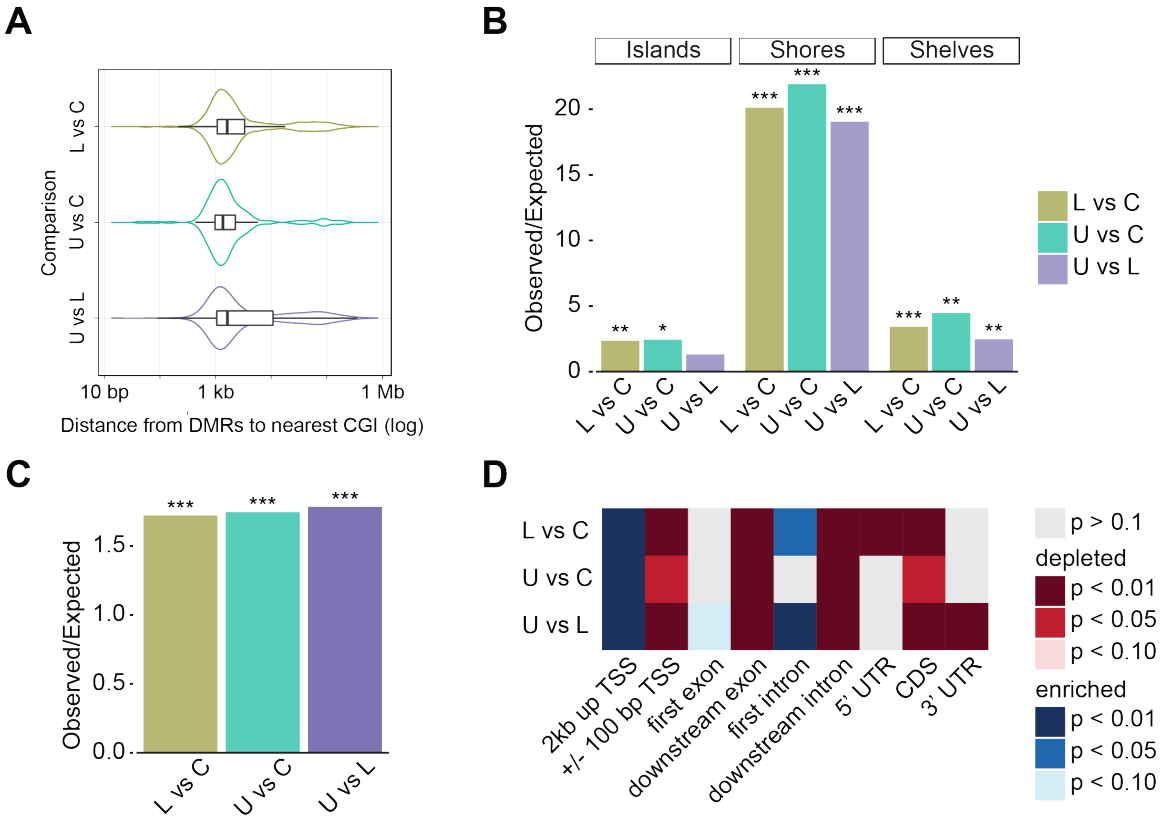


Figure 4.4 Spatial correlation between DMRs and CGI or gene-associated features. (A–C) Lower vs. control (L vs C), upper vs. control (U vs C), and upper vs. lower (U vs L) comparisons. **(A)** Violin and box plots showing distribution of distance from DMRs to center of nearest CGI. **(B–C)** Binomial enrichment of DMRs at CpG Islands (CGIs), CGI shores, CGI shelves **(B)**, and gene-associated regions (e.g., promoter or gene body) (C) (*p < 0.05, **p < 0.01, ***p < 1 x 10^{-10}). **(D)** Overlap of DMRs at genic features within gene-associated regions calculated by Jaccard index permutation test.

3. Discussion

Accumulating evidence has demonstrated that adult onset metabolic diseases, including obesity, diabetes, and cardiovascular disease, are linked to reprogramming events during fetal development. Importantly, the incidence of these disorders has arisen during just the past few decades and increasing data have shown that the trend is attributed to environmental factors (Thayer et al., 2012). Because the timeline of disease occurrence overlaps with the introduction and increasing use of man-made chemicals in our environment, one hypothesis is that exposure to these compounds alters fetal reprogramming events and predisposes individuals to metabolic disease. Despite support from animal and human studies, the mechanisms linking exposure to the phenotypes are unclear, and furthermore, it is unknown whether different environmental factors act through distinct mechanisms or whether core mechanisms exist. Elucidating the relevant mechanisms will provide important insights into risk management and disease prevention for the protection of the health of our future generations.

The goal of the current study is to elucidate pathways that underlie metabolic abnormalities induced by early life exposure to environmental chemicals using the mouse model. Previously, our laboratory had demonstrated that early life exposure to BPA induced multigenerational, male-specific perturbations of metabolic health in mice (Susiarjo et al., 2015). In the study, we have used a combined -omics approach to determine how transcriptional and epigenetic changes are linked to altered metabolome in BPA-exposed mice. We selected BPA as a representative environmental chemical due to its widespread presence in the environment and because most people are exposed to this compound. Furthermore, because many of its characteristics are shared with other environmental contaminants (McLachlan, 2016) that have also been linked to developmental abnormalities in humans and animals, elucidating molecular mechanisms relevant to BPA exposure can provide general insights into how xenobiotics impact our health.

DNA methylation-independent pathway perturbation

Changes in mRNA levels do not necessarily require substantial alterations to DNA methylation changes. In addition to pathways associated with DMRs, we observed pathways that were not directly associated with DNA methylation changes. Through the utilization of the multi -omics approach, we have

discovered some potentially important metabolic pathways relevant to environmental exposure as modeled by BPA. Metabolomics, for example, revealed that *in utero* BPA exposure resulted in misregulation of lipid metabolism (Susiarjo et al., 2017). Specifically in the upper dose exposure group, BPA exposure upregulated levels of primary and secondary bile acids (Susiarjo et al., 2017). Bile acids are "physiologically detergent" molecules that facilitate the absorption of dietary lipids and vitamins in the gut (Taoka et al., 2016). Synthesized from cholesterol in the liver, bile acids are signaling molecules that regulate lipid, glucose, and energy metabolism through the action of the bile acid-activated nuclear receptor, the Farnesoid X Receptor or FXR (Taoka et al., 2016). Increasing evidence has demonstrated that perturbed bile acid function is linked to altered glucose homeostasis and other metabolic abnormalities. Consistent with these observations, IPA using our RNA-seq data implicated the FXR activation and cholesterol biosynthesis (Supplemental Tables S4.4, S4.6) as pathways perturbed in upper dose BPA exposed-mice, suggesting that BPA exposure-induced metabolic abnormalities are linked to misregulation of genes relevant to bile acid and/or cholesterol synthesis. These findings suggest that bile acid metabolism could represent a common pathway implicated in environmental exposure-induced metabolic changes.

Although we did not identify perturbed FXR pathways in our DNA methylation data, upper dose fetal livers had promoter hypermethylation and decreased expression of *Tcf4* (Transcription factor 4) compared to lower dose fetal livers. *Tcf4* is a basic helix-loop-helix (bHLH) transcription factor that has recently been identified as a risk locus for primary sclerosing cholangitis in humans (Ellinghaus et al., 2013; Forrest et al., 2014). Primary sclerosing cholangitis is characterized by inflammation of hepatic bile ducts, which can lead to accumulation of bile in the liver (Forrest et al., 2014; Li et al., 2017). Our data suggests a possible indirect link between BPA-induced, DNA methylation-associated gene expression change of *Tcf4* and hepatic bile acid dysregulation. Additional studies are needed to determine downstream targets of TCF4-mediated signaling and whether bile acid abnormalities precede or result from altered *Tcf4* expression.

DNA methylation-associated pathway perturbation

BPA altered DNA methylation and mRNA expression of nutrient sensing genes, free fatty acid receptor 4 (*Ffar4*) and protein kinase C-zeta (*Prkcz*). FFAR4 (GPR120) is an omega-3 fatty acid (ω-3 FA)

receptor that plays a role in lipid sensing and mediates protective effects from dietary ω-3 FAs (Da Young Oh et al., 2010; Ichimura et al., 2012). WGBS data showed increased methylation in lower (66.8%) and upper (65.9%) dose group compared to control (49.2%) in the first intron (chr19:38098849-38099269; 6 CpG sites) of *Ffar4*. Lower dose had significant decreased expression (-2.47 log2 fold-change; FDR = 0.02) and upper dose and had moderate, but non-significant, decreased expression (-0.11 log2 fold-change). *Prkcz* is involved in insulin signaling (Chen et al., 2015) and has two transcript variants, determined by alternate promoter usage, that encode for atypical protein kinase C (PKC) enzymes: the first variant encodes PKCζ with a catalytic domain and an autoinhibitory regulatory domain; the second variant lacks the regulatory domain (Hernandez et al., 2011). For lower vs. control comparison, we identified a DMR (15.9% increased methylation) at the promoter of the transcript variant that encodes for the constitutively active isoform (chr4:155346128-155347203; 23 CpG sites; fourth intron of transcript for autoinhibited PKCζ). Although not retained after lfdr correction, we also observed increased methylation in upper (51.5%) compared to control (43.1%) fetal livers. *Prkcz* had significantly increased mRNA expression (2.32 log2 fold-change, FDR = 0.03) in lower and moderate, but non-significant, increased mRNA expression (0.3 log2 fold-change) in upper compared to control fetal livers. DMR location suggests potential for transcript-specific regulation of *Prkcz*. Additional studies are needed to determine whether increased methylation is suppressing transcription of the constitutively active isoform while increasing expression of the autoinhibited isoform or increasing transcription of both isoforms. Taken together, these data demonstrate DNA methylation-mediated alterations to nutrient sensing in BPA exposed mice with potential implications for metabolic health.

Furthermore, metabolomics data showed significant upregulation of glutamate in the upper dose relative to lower (Susiarjo et al., 2017). Recent studies have demonstrated the role of cytosolic glutamate as a key molecule linking glucose metabolism to incretin/cAMP action to amplify insulin secretion (Gheni et al., 2014). Interestingly, previously we found that F1 males exposed to both upper and lower dose BPA during *in utero* development developed glucose intolerance (Susiarjo et al., 2015). F1 males exposed to upper dose BPA were insulin resistant, but their islets exhibited normal glucose-stimulated insulin release, in contrast to islets isolated from males in the lower dose that showed significantly reduced glucose-stimulated insulin release. Consistent with these observations, IPA using both RNA-seq and WGBS data

revealed glutamate receptor signaling to be one of the pathways that was differentially regulated in the lower vs. upper dose exposure groups, suggesting that either changes in glutamate receptor signaling are linked to insulin secretory defects in the lower dose group and/or they have protective effects on islets from males exposed to upper dose BPA. These observations suggest that multiple approaches are required to elucidate mechanistic pathways underlying metabolic diseases induced by environmental exposure. Moreover, future works should include investigation of glutamate signaling in environmental exposure-induced perturbed glucose homeostasis.

Greater number of DEGs and DMRs at lower dose than upper dose

The RNA-seq and WGBS data revealed significantly more DEGs and DMRs in livers from lower dose BPA exposure relative to upper dose. This could be due to the observation that the upper dose produced significantly more substantial changes in metabolism (Susiarjo et al., 2017). We previously reported that F1 male mice exposed to upper dose BPA had increased body fat and developed glucose intolerance and insulin resistance as adults, while the phenotype of the lower dose-exposed mice was milder (Susiarjo et al., 2015). Additionally, lower BPA dose exposure may produce more complex changes relative to upper dose, changes that involved both non-metabolic and metabolic perturbations as shown by the DEGs and DMRs. Alternatively, the inconsistencies could originate from technical limitations and may be explained by the restricted number of major metabolite classes analyzed by the Metabolon DiscoveryHD4 platform used in the metabolomics study (Susiarjo et al., 2017). This limitation contrasts with the less biased techniques of RNA-seq and WGBS. Furthermore, changes of the transcriptome and/or methylome may not immediately affect the metabolome.

Heterogeneity in response to lower dose exposure

Intriguingly, gene expression and methylation studies revealed that BPA exposure, especially the lower dose, produced heterogeneous responses among genetically identical mice. In the WGBS, our data clustered mice from the lower dose exposure group into two distinct groups: one group with somewhat similar DNA methylation pattern to control and upper dose fetal livers and a second group distinct from both control and upper dose fetal livers. These data suggest that BPA-induced DNA methylation changes

may follow a bimodal (or multimodal) distribution. Notably, multimodal shifts in methylation patterns from lower dose exposure provide an attractive explanation of observed bimodal shift in distribution of coat colors in A^{vy}/a offspring at lower dose BPA (Anderson et al., 2012). Using a developmental exposure model with coat color of A^{vy}/a offspring as a biosensor of nutritional and environmental exposure-induced epigenetic perturbation (Dolinoy, 2008) reviewed in (Martos et al., 2015), Anderson and coworkers demonstrated that coat color displayed a bimodal response to lower BPA exposure (Anderson et al., 2012). That is, they showed decreased proportions of mottled (the modal value for phenotype in controls) and increased proportions of both slightly mottled (phenotype indicative of decreased methylation) and pseudoagouti (phenotype indicative of increased methylation). Combined with a higher standard deviation in methylation at the A^{vy} locus at lower dose relative to controls reported in the study (Anderson et al., 2012), the changes in coat color phenotype support the genome-wide results from our WGBS data that lower dose BPA produces heterogeneity in DNA methylation at some loci. While further studies with more biological replicates at targeted loci are needed to characterize the bimodal (or multimodal) response, the observed heterogeneity in response to lower dose BPA exposure provides a unique model for exploring interindividual variability in response to environmental exposures.

Concordance with previous studies

To our knowledge, this is the first study that employs multiple-omics with the goal of dissecting DNA methylation-mediated and DNA methylation-independent pathways related to EDC exposure. A study conducted by van Esterik and coworkers investigated solely the genome wide DNA methylation changes in adult liver of C57BL/6JxFVB hybrid mice that were exposed to 3000 µg/kg bw of BPA during gestation and lactation via maternal diet. In this model system, they had found female-specific obesity and metabolic impairment (van Esterik et al., 2014). Through DREAM, the authors identified 19 sites with at least 10% DNA methylation differences and unadjusted $p < 10^{-2}$ (van Esterik et al., 2015). Upon comparison with our WGBS data, we observed that 10 out of the 19 genes were also differentially methylated in our study prior to false discovery rate correction: 4 DMRs exclusively in the lower dose (*Evpl*, *Col4a4*, *Scube3*, and *Tcf3*), 1 in the upper dose group (*F8a*), and 5 that were differentially methylated in both the upper and lower exposure groups (*Chordc1*, *Dpt*, *Pcdh15*, *Atp2b2*, and *Taf7l*). In

Iqbal et al (2015), the authors studied the potential transgenerational inheritance of altered epigenome in JF1XOG2 hybrid mice exposed to EDCs including 100 mg/kg/day vinclozolin and 200 µg/kg/day BPA. They studied altered DNA methylation at imprinted and non-imprinted loci in various embryonic tissues, including the liver. Altered methylation of the *Dlk1* and *Ascl2* genes (Iqbal et al., 2015) were also observed in our current study. While our results have confirmed previous findings, the multi-omics approach has revealed considerably more information.

Study Limitations

Although we have identified methylation-associated and methylation-independent pathway perturbations from prenatal exposure to BPA, our data is from a single developmental stage. This makes integrating the data in a way that informs causal inference into molecular mechanisms difficult because we cannot determine the temporal dependence of the observed differences among DNA methylation and gene expression data. In other words, from a single time point we cannot differentiate between whether the observed DNA methylation changes result in or are the result of changes in gene expression. We can, however, conclude that these coinciding changes precede the metabolic phenotypes observed in adult animals. As such, we have provided unprecedented insight into physiologically relevant pathways that have potential importance for adult and multigenerational phenotypes.

Additionally, changes observed in exposed fetuses could be transient and return to baseline once the stimulus (i.e., BPA exposure) is removed. Importantly, a transient change could still result in altered fetal programming if the temporary change initiates a cascade of events that eventually lead to an altered physiological outcome. However, the details for these mechanisms are beyond the scope of this work. To determine the extent to which these pathways are causally involved in fetal reprogramming, future time-course studies in postnatal through the adult life stages are needed to follow the DNA methylation and gene expression changes reported here.

4. Conclusions

In summary, we have demonstrated that *in utero* exposure to BPA significantly altered methylome and transcriptome of the fetus in a dose-dependent manner. Integration of methylation and gene expression

data indicates a potential role for DNA-methylation mediated perturbation of RICTOR/MTORC2, SCAP/SREBP, nutrient sensing, and glutamate pathways. In addition, we observed that many changes occurred in pathways related to lipid metabolism, specifically cholesterol biosynthesis and bile acid regulation that were not directly associated with DNA methylation changes. The multi-omics approach used in the current study is a powerful method to elucidate pathways disrupted by BPA that could provide biomarkers to predict future disease susceptibility and potential therapies. Our findings suggest that DNA methylation-mediated and DNA methylation-independent gene expression changes likely work simultaneously to culminate in altered fetal programming in response to environmental exposures. As such, future mechanistic studies should focus on how both altered signaling pathways and epigenetic marks coordinate to regulate these pathways in metabolic disorders.

5. Materials and Methods

Mouse information

Six weeks old virgin C57BL/6J female mice were exposed to control, 10 μg/kg bw of BPA (lower dose), and 10 mg/kg/bw of BPA (upper dose) through diet (Envigo Inc, Madison, WI) two weeks prior to mating, and exposure continued during mating until E18.5. At E18.5, pregnant mice were sacrificed, and we isolated liver tissues from the fetuses. The sex of the fetuses was determined by physical appearance of the gonads, and confirmed by sex-specific PCR. Liver tissues from both males and females were weighed, snap frozen, and stored at –80° C until analysis. We previously reported male-specific metabolic phenotypes (Susiarjo et al., 2015), so therefore, although we collected both male and female fetal tissues in this work, only male fetuses were included in the RNA-seq and WGBS studies.

RNA sequencing

Total RNA was extracted from mouse fetal liver tissues using the RNeasy Mini Kit (Qiagen) and quantified using a NanoDrop spectrophotometer. Approximately 10 μg of total RNA was used to isolate polyadenylated (polyA) using the Dynabeads mRNA DIRECTTM Purification kit (Thermo Fisher Scientific, Waltham, MA). RNA-seq library preparation was prepared using the ScriptSeq v2 technology (Illumina, San Diego, CA) according to manufacturer protocol. Briefly, approximately 50 ng of poly(A)+

RNA was fragmented and reverse transcribed using random primers containing a 5'-tagging sequence. The 5'-tagged cDNA was then tagged at its 3' end by the terminal-tagging reaction to yield di-tagged, single-stranded cDNA. Following purification, the di-tagged cDNA was amplified by limited cycle PCR, which completed the addition of the Illumina adaptor sequences, amplified the library for subsequent cluster generation, and added Illumina index barcodes. The amplified RNA-seq library was analyzed using qPCR (Kapa Biosystems, Wilmington, MA) and BioAnalyzer (Agilent Technologies, Santa Clara, CA) for quantity and quality assessment, respectively. Equal moles of each library were pooled and re-quantified by qPCR (Kapa Biosystems, Wilmington, MA) and sequenced using the NextSeq 500 platform (Illumina, San Diego, CA).

As quality control, paired-end reads were trimmed using Trimmomatic (Bolger et al., 2014) with a mild trimming criterion (Phred < 3), and assessed using FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). Concordant paired reads with good quality were mapped to reference genome (mm10) using STAR (Dobin et al., 2013) (mapping quality >=10), and assessed using RSeQC (Wang et al., 2012). The expression level of genes was quantified using HTSeq (Anders et al., 2015) and DESeq2 (Love et al., 2014) was used to determine differentially expressed genes (false discovery rate < 0.05).

Whole genome bisulfite sequencing

WGBS libraries were constructed as described in (Li et al., 2015). Genomic DNA (250 ng - 1 μ**g**) was sheared to 200-500 bp with Bioruptor Plus sonication device (B01020001, Diagnode, Denville, NJ). After purification with Agencourt Ampure XP (A63881, Beckman Coulter, Brea, CA), end repair of sheared DNA was performed with End-It DNA End Repair Kit (ER81050, Epicentre, Madison, WI). DNA was purified with Agencourt Ampure XP and adenine was added with 10 ***μ***l of 1 mM dATP and 3 ***μ***l Taq DNA polymerase (M0267S, New England Biolabs [NEB], Ipswich, MA) in a 50 ***μ***l reaction, incubated at 72°C for 30 minutes. Following purification with Agencourt Ampure XP, Illumina Truseq adapters were ligated to adenylated DNA fragments with 4 ***μ***l T4 DNA ligase (M0202L, NEB, Ipswich, MA) in a 40 ***μ***l reaction, incubated for 30 minutes at room temperature and then 16^{o}C overnight. Adapter-ligated DNA fragments were size selected to 350-600 bp with 2% agarose gel and MinElute Gel Extraction Kit (28604,

Qiagen, Hilden, Germany). Size-selected fragments were bisulfite treated with Imprint DNA Modification Kit (MOD50-1, Sigma-Aldrich, St. Louis, MO) and PCR amplified (12-14 cycles) with KAPA HiFi HotStart Uracil+ ReadyMix (KK2802, Kappa Biosystems, Wilmington, MA). Amplified libraries were gel purified with 2% agarose gel and MinElute Gel Extraction Kit. Qubit dsDNA HS Assay (Q32854, Life Technologies, Eugene, OR) and High Sensitivity DNA Kit for Agilent 2100 Bioanalyzer (5067-4626, Agilent Technologies, Waldbronn, Germany) were used to determine library concentration and size, respectively. Beckman Coulter Genomics performed paired-end sequencing (2 x 100 bp or 2 x 125 bp; Danvers, MA).

Sequencing reads were trimmed with Trim Galore version 0.4.0 (www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using Cutadapt version 1.8.1 (Martin, 2011) and FastQC version 0.11.2 (www.bioinformatics.babraham.ac.uk/projects/fastqc/) with options --paired --three_prime_clip_R1 3 --three_prime_clip_R2 3 --clip_R2 2. Trimmed reads were aligned to GRCm38/mm10 genome with Bismark version 0.14.5 implementing Bowtie2 version 2.2.5 (Langmead et al., 2009) and options -N 1 -D 20 -R 5 -X 1200 --chunkmbs 1024. PCR duplicates were removed from aligned pairs with deduplicate_bismark and methylation status of cytosines was determined by bismark_methylation_extractor, ignoring calls in the first six base pairs of each read to prevent methylation bias from the end repair step of library construction (Krueger and Andrews, 2011). CpG methylation calls from opposite strands were merged into a single call per CpG (Krueger and Andrews, 2011).

Identification of Differentially Methylated Regions and Local False Discovery Rate

Post-processing analyses were carried out in R version 3.2.2. Technical replicates were merged and CpG sites covered by more than 500 reads were removed to eliminate regions with potential PCR bias (Akalin et al., 2012). Smoothed methylation estimates were determined using a local-likelihood method and modified t-statistics were calculated for single CpG sites that had at least 2X coverage for at least 2 replicates in each exposure group in the comparison. Differentially methylated regions (DMRs) were constructed by combining consecutive CpG sites in the lower 2.5th and upper 97.5th percentiles into hypomethylated and hypermethylated DMRs, respectively, with the bsseq R package (Hansen et al., 2012). To determine the local false discovery rate (lfdr), we applied empirical null hypothesis modeling to the

single-CpG corrected t-statistics (Efron, 2007). After transforming t-statistics into z-scores (Efron, 2007), we calculated p-values, lfdr, and q-values (tail-area based false discovery rate) using the fdrtool (Strimmer, 2008) R package (Supplemental Fig. S4.2). Correlation among CpG sites violates the independence requirement of tail-area-based multiple hypotheses testing correction methods; therefore, we used empirical null modeling to control the false discovery rate (FDR) of DMRs (Efron, 2007). DMRs containing at least one CpG with an lfdr less than 0.2 were retained.

Hierarchical Clustering, Annotation, and Enrichment Analyses

For hierarchical clustering, DMRs for each pair-wise comparison (i.e., control vs. lower dose, control vs. upper dose, and lower vs. upper dose) were merged to generate a set of non-overlapping DMRs. Average methylation across each merged DMR was calculated for individual biological replicates. Samples were clustered by Ward's method with Euclidean distances and visualized as heatmap using NMF R package (Gaujoux and Seoighe, 2010). For annotation, DMRs were assigned to promoters, gene body, or intergenic with annotations from TxDb.Mmusculus.UCSC.mm10.knownGene (R package) based on the location of the DMR midpoint. We defined promoters as the region between 2 kb upstream and 1 kb downstream of a transcription start site (TSS), and gene body as 1 kb downstream from TSS to the transcription end site (TES). Genome-wide enrichment at CpG islands (CGIs), CGI shores, CGI shelves, and gene-associated regions was determined by two-sided binomial test. CGIs were downloaded from UCSC mm10 unmasked CpG Islands track. CGI shores are defined as 2 kb regions flanking CGIs that do not overlap a CGI (Irizarry et al., 2009) and CGI shelves are defined as 2 kb regions flanking CGI shores that do not overlap a CGI or CGI shelf (Bibikova et al., 2008). We defined gene-associated regions as 2 kb upstream from a TSS to TES (i.e., containing promoter and gene body of a transcript). Distance from CGI midpoint to DMR midpoint and absolute distance permutation test (100 permutations) was performed with GenometriCorr R package (Favorov et al., 2012). Overlap of DMRs at genic features within gene-associated regions was calculated by Jaccard index and significance of overlap was determined by permutation test. We identified genic features—within 2 kb upstream of TSS (-2 kb TSS), within 100 bp upstream or downstream from TSS (+/- 100 bp TSS), first exon, downstream exon, first intron, downstream intron, 5' untranslated region (UTR), 3' UTR, and coding DNA sequence (CDS) exon—for mm10

transcripts with annotations from TxDb.Mmusculus.UCSC.mm10.knownGene (R package). Null distributions for permutation tests were constructed by randomly distributing each DMR along the overlapping genic region and calculating Jaccard index for a specific genic feature (100 permutations). Genic regions for Jaccard permutation tests were defined as 2kb upstream from TSS to TES, with regions containing multiple overlapping genes or transcripts merged into a single genic region. Binomial and Jaccard permutation tests were performed with GenometriCorr R package (Favorov et al., 2012).

Pathway Analyses and Gene Set Enrichment for Transcription Factor Binding Targets

For RNA-seq and/or WGBS studies, pathway and functional analyses were generated through the use of QIAGEN's Ingenuity Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity). MSigDB version 5.1 was used to identify transcription factors and overrepresented Transcription Factor Binding Site (TFBS) in promoters of genes with promoter DMRs with the Transcription Factor Targets gene set and FDR less than 0.05 (Subramanian et al., 2005).

6. List of Supplemental Materials

Supplemental Figures S4.1 to S4.2 (**Appendix III**)

- Figure S4.1 Genome-wide coverage and methylation for WGBS samples.
- Figure S4.2 Distribution of statistics for single-CpG methylation differences in BPA-exposed male fetal livers.

Supplemental Tables S4.1 to S4.12

- Table S4.1 Alignment metrics for RNA-seq data
- Table S4.2 Differentially expressed genes (DEGs)
- Table S4.3 Ingenuity Pathway Analysis for RNA-seq DEGs: Lower vs. Control
- Table S4.4 Ingenuity Pathway Analysis for RNA-seq DEGs: Upper vs. Control
- Table S4.5 Ingenuity Pathway Analysis for RNA-seq DEGs: Upper vs. Lower
- Table S4.6 Ingenuity Pathway Analysis for RNA-seq DEGs: Top Pathways
- Table S4.7 Differentially Methylated Regions (DMRs)
- Table S4.8 Gene expression levels at DMRs
- Table S4.9 Ingenuity Pathway Analysis for WGBS DMRs: Lower vs. Control
- Table S4.10 Ingenuity Pathway Analysis for WGBS DMRs: Upper vs. Control
- Table S4.11 Ingenuity Pathway Analysis for WGBS DMRs: Upper vs. Lower
- Table S4.12 Gene set enrichment for Transcription Factor Binding Sites: Upper vs. Lower

Chapter 5

Conclusions and Future Directions

The overall goal of this work was to increase the understanding of how environmental stressors may alter epigenetic states by identifying and characterizing genomic features that are susceptible to epigenetic perturbation during early development. Our perspective is that better knowledge of the molecular basis for how epigenetic marks, such as DNA methylation, are established and maintained throughout the mammalian genome will contribute to identifying the role that epigenetic mechanisms play in human disease, specifically diseases with an environmental component. Ultimately, determining how the heritability and plasticity of epigenetic states contribute to mechanisms of toxicity from intergenerational and transgenerational exposures will allow us to distinguish between biomarkers of exposure (e.g., correlated genome-wide changes) and causal molecular pathways that contribute to disease pathology.

In the first part of Chapter 1, we reviewed literature on epigenetic inheritance across multiple generations. In contrast to non-mammalian eukaryotes, the mammalian life cycle reduces the likelihood that environment-induced epigenetic changes would be inherited, independent from genetic inheritance, through the germline. Nonetheless, transgenerational silencing of transposable elements and selective parent-of-origin-dependent regulation of genomic imprints serve as examples of epigenetic inheritance across multiple generations in mammals. We explained the inherent difficulties in identifying multigenerational phenotypes and in connecting molecular changes to disease states in humans. Knowledge of molecular events linking prenatal exposures to adult phenotypes through somatic inheritance of environment-induced epigenetic changes in rodent models is also limited. Furthermore, interspecies differences in genomic features associated with epigenetic inheritance (e.g., retrotransposons and imprints) limits the direct extrapolation of phenotypes observed in mice to humans. We proposed that a molecular understanding of epigenetic inheritance in animal models would allow for mechanistic extrapolation to identify the analogous features within the human genome. That is, if we can identify the molecular mediators acting on mouse genomic elements, then we could discover genomic elements (e.g., genes) affected by those mediators in humans and explore the association between candidate human genes and disease etiology.

To that end, we previously characterized the extent to which mouse genomic features relied on different DNA methlytransferase (DNMT) enzymes, DNMT1 or DNMT3a/3b to retain their methylation in

mouse embryonic stem cells (mESCs) (Li et al., 2015). Those findings led to an updated coordination and division of labor model for DNA methylation maintenance, described in the second part of Chapter 1. In the updated model, we proposed that the majority of CpG sites rely on the coordination of enzymatic activities from DNMT1 and DNMT3a or 3b. However, a subset of genomic regions, which retained high methylation in the absence of DNMT3a/3b, was considered to have DNMT1-dependent maintenance. Similarly, regions that were resistant to methylation loss in the absence of DNMT1 were considered to have DNMT3a/3b-dependent maintenance.

In this work, we sought to determine whether certain genomic regions displayed differential vulnerability to epigenetic perturbation. In Chapter 2, we began with a genetic knockout model, using DNMT1-deficient mESCs with exogenous *Dnmt1* cDNA expression. In this genetic model of epigenetic perturbation, we described how germline imprinted regions were susceptible to non-restorable DNA methylation loss, despite genome-wide increase in methylation levels in "rescued" DNMT1-deficient mESCs compared to DNMT1-deficient mESCs.

We developed a bioinformatic approach to identify 2500 regions throughout the genome that had **no**n-**re**scued **D**MRs (NORED). We demonstrated that NORED is a novel and robust method. We then applied a second bioinformatic approach, MethylMosaic, to quantify read-level (as opposed to CpG site-level) methylation genome-wide to identify 2500 genomic regions with bimodal methylation patterns. This characteristic DNA methylation pattern emerges at imprinted loci in bisulfite sequencing reads–where one read represents a single DNA molecule–from one fully methylated and one unmethylated allele. A systematic comparison of regions detected by NORED and MethylMosaic demonstrated that these two approaches are complementary, but independent methods for identifying distinct DNA methylation characteristics. Use of reciprocal hybrid cross embryonic stem cells, which allowed us to distinguish between parent-of-origin and genetic influences on allelic methylation, demonstrated that regions with epigenetic properties in common with imprinted genes are not necessarily imprinted.

Taken together, results reported point to several paradigm shifts. First, monoallelic methylation and DNMT1-dependence are distinct epigenetic characteristics and both occur at imprinted gDMRs. Second, having both bimodal methylation and DNMT1-dependence, is not restricted to imprinted gDMRs.

Including regions at 19 established imprinted gDMRs, we identified 207 regions with at least 123 potentially associated genes that were detected by both methods. Experimental data demonstrate that at least two of these loci, *Hcn2* (Chapter 2, Fig. 2.4) and *Park7* (Chapter 2, Fig. 2.5), have bimodal methylation patterns that are parent-of-origin independent in hybrid embryonic stem cells. This new evidence for probabilistic regulation of bimodal methylation patterns leads to the novel concept of 'switchable' allelic methylation, reminiscent of random X-chromosome inactivation (XCI).

Importantly, distinguishing between three groups–exclusively bimodal, exclusively DNMT1-dependent, and both bimodal and DNMT1-dependent–defined by combining the two methods has functional implications for common diseases. The separation into groups is particularly informative for mechanistic studies in pharmacology, nutrition, and toxicology. Monoallelically expressed genes could be considered susceptible to a single hit from a mutagen (or drug) because there is not a second expressed allele to compensate. DNMT1-dependent regions could be targets for DNMT inhibitors or exposures affecting methyl donors (e.g., nutrition and arsenic), working through epigenetic mechanisms.

Notably, regions identified by both approaches are expected to exhibit vulnerability to both genetic and epigenetic drugs/toxicants/nutritional exposures. While monoallelic expression is recognized as a rationale for the susceptibility of imprinted genes to mutagenic stressors or *de novo* mutations, the additional layer of DNMT1 dependence represents a fundamental shift in the underlying assumptions about the inherent vulnerability of imprinted gDMRs. Furthermore, the identification of non-imprinted regions that share these two attributes with imprinted gDMRs, greatly expands the pool of candidate genes for 'developmental origin of adult disease' and 'transgenerational epigenetic inheritance.' We envision that our characterization of DNMT1-dependent regions and regions with bimodal methylation patterns will be informative for analyses of methylation differences in response to toxicant/drug exposure.

In Chapter 3 we sought to determine whether DNA methylation characteristics associated with gDMRs and certain classes of transposable elements increased susceptibility to epigenetic perturbation from prenatal exposure to Bisphenol A (BPA). We found that CpG sites within regions with imprint-like DNA methylation properties (bimodal, bimodal NORED, and NORED) or distinct DNA methylation maintenance paradigms (DNMT1 maintained or DNMT3a/3b maintained) had increased likelihood for

BPA-induced DNA methylation changes compared to CpG sites genome-wide. Bimodal NORED regions exhibited the most striking effect and included altered CpG sites from both known imprinted loci and loci not identified as imprinted. DNMT1-maintained regions had the lowest correlations in CpG site- and region-level methylation between dose groups. Characterization of retrotransposons indicated considerable differences in response to BPA among subtypes of ERVK family LTR retrotransposons. Several ERVK family subtypes that displayed the lowest correlations in region-level methylation between dose groups had been previously identified as highly DNMT1 dependent (Li et al., 2015), including IAPLTR1 Mm elements.

In Chapter 4 we aimed to explore the functional relevance of altered DNA methylation from prenatal BPA exposure. We identified dose-dependent, BPA-induced DNA methylation changes in several genes with potential consequences for metabolic health. The broad perspective on DNA methylation change-associated functional pathways presented here illuminated prospective molecular mediators of adult, and possibly intergenerational, metabolic phenotypes. However, detailed mechanistic studies are needed to establish a role for these changes in disease etiology. Specifically, studies should be conducted to examine potential pathways across multiple time points throughout the life course of the animals.

Future studies should explore the relationship between intraspecies variability in site-specific methylation and DNMT1-dependent DNA methylation. Our work raises the possibility that DNMT1-maintained regions could be more variable due to inefficiency of DNMT1 leading to probabilistic loss of methylation. This also raises the question of whether environmental exposure-induced shifts (e.g., BPA) in methylation level at these loci is more likely because BPA-altered methylation is mediated by specific targeting of DNMT1, either directly or indirectly through signaling pathways leading to transcription factor recruitment toward or sequestration from certain genomic regions. This would have implications for identifying functional consequences of loci with environment-influenced changes in DNA methylation. This work indicates that environmental stressors working through epigenetic mechanisms may exhibit probabilistic dose-responses if the genes contributing to the phenotypes of interest rely predominantly on DNMT1 for methylation maintenance.

Finally, our present work focused almost exclusively on DNA methylation. Other epigenetic marks, such as a wide variety of histone modifications, play a role in epigenetic regulation of genomic loci. As such other epigenetic mechanisms are likely to be important in determining region- and locus-specific susceptibility to epigenetic perturbation. It is possible that DNA methylation changes described here are representative of the epigenetic state at a genomic locus, but not necessarily the causal molecular driver for activation or silencing of gene or repeat element transcription. However, the internal normalization of bisulfite sequencing is more conducive to making site-specific comparisons across samples than count-based, genome-wide methods. For count-based methods it is impossible to distinguish between lack of coverage at a site and zero signal, whereas with WGBS it is possible to quantify zero methylation at a covered CpG site. As such, DNA methylation changes are an indicator of epigenetic perturbation. Therefore future studies should aim to understand how the interaction between DNA methylation and other epigenetic marks (e.g., histone modifications) contribute to regional susceptibility to epigenetic perturbation at loci identified in this work.

References

Acevedo, N., Davis, B., Schaeberle, C.M., Sonnenschein, C., and Soto, A.M. (2013). Perinatally Administered Bisphenol A Acts as a Mammary Gland Carcinogen in Rats. Environ Health Perspect 1–7.

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L., et al. (2012). Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. PLoS Genet. *8*, e1002781–16.

Alonso-Magdalena, P., Quesada, I., and Nadal, A. (2015). Prenatal Exposure to BPA and Offspring Outcomes. Dose-Response *13*, 155932581559039–8.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anderson, O.S., Nahar, M.S., Faulk, C., Jones, T.R., Liao, C., Kannan, K., Weinhouse, C., Rozek, L.S., and Dolinoy, D.C. (2012). Epigenetic responses following maternal dietary exposure to physiologically relevant levels of bisphenol A. Environ. Mol. Mutagen. *53*, 334–342.

Angle, B.M., Do, R.P., Ponzi, D., Stahlhut, R.W., Drury, B.E., Nagel, S.C., Welshons, W.V., Besch-Williford, C.L., Palanza, P., Parmigiani, S., et al. (2013). Metabolic disruption in male mice due to fetal exposure to low but not high doses of bisphenol A (BPA): Evidence for effects on body weight, food intake, adipocytes, leptin, adiponectin, insulin and glucose regulation. Reproductive Toxicology *42*, 256–268.

Anway, M.D., Memon, M.A., Uzumcu, M., and Skinner, M.K. (2006a). Transgenerational Effect of the Endocrine Disruptor Vinclozolin on Male Spermatogenesis. Journal of Andrology *27*, 868–879.

Anway, M.D., and Skinner, M.K. (2008). Transgenerational effects of the endocrine disruptor vinclozolin on the prostate transcriptome and adult onset disease. Prostate *68*, 517–529.

Anway, M.D., Cupp, A.S., Uzumcu, M., and Skinner, M.K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. Science *308*, 1466–1469.

Anway, M.D., Leathers, C., and Skinner, M.K. (2006b). Endocrine Disruptor Vinclozolin Induced Epigenetic Transgenerational Adult-Onset Disease. Endocrinology *147*, 5515–5523.

Anway, M.D., Rekow, S.S., and Skinner, M.K. (2008). Transgenerational epigenetic programming of the embryonic testis transcriptome. Genomics *91*, 30–40.

Arnaud, P. (2010). Genomic imprinting in germ cells: imprints are under control. Reproduction *140*, 411–423.

Babak, T., DeVeale, B., Tsang, E.K., Zhou, Y., Li, X., Smith, K.S., Kukurba, K.R., Zhang, R., Li, J.B., van der Kooy, D., et al. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat. Genet. *47*, 544–549.

Bailey, K.A., and Fry, R.C. (2014). Arsenic-Associated Changes to the Epigenome: What Are the Functional Consequences? Curr Envir Health Rpt *1*, 22–34.

Barbaux, S., Gascoin-Lachambre, G., Buffat, C., Monnier, P., Mondon, F., Tonanny, M.-B., Pinard, A., Auer, J., Bessières, B., Barlier, A., et al. (2012). A genome-wide approach reveals novel imprinted genes expressed in the human placenta. Epigenetics *7*, 1079–1090.

Barker, D.J., and Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in England and Wales. Lancet *1*, 1077–1081.

Barker, D.J., Osmond, C., Simmonds, S.J., and Wield, G.A. (1993). The relation of small head circumference and thinness at birth to death from cardiovascular disease in adult life. Bmj *306*, 422–426.

Barouki, R., Gluckman, P.D., Grandjean, P., Hanson, M., and Heindel, J.J. (2012). Developmental origins of non-communicable disease: implications for research and public health. Environ Health *11*, 42.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837.

Bartolomei, M.S., and Ferguson-Smith, A.C. (2011). Mammalian genomic imprinting. Cold Spring Harbor Perspectives in Biology *3*.

Bateson, P., Barker, D., Clutton-Brock, T., Deb, D., D'Udine, B., Foley, R.A., Gluckman, P., Godfrey, K., Kirkwood, T., Lahr, M.M., et al. (2004). Developmental plasticity and human health. Nature *430*, 419–421.

Bestor, T.H., and Ingram, V.M. (1983). Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci U S A *80*, 5559–5563.

Bibikova, M., Laurent, L.C., Ren, B., Loring, J.F., and Fan, J.-B. (2008). Unraveling epigenetic regulation in embryonic stem cells. Cell Stem Cell *2*, 123–134.

Bindhumol, V., Chitra, K.C., and Mathur, P.P. (2003). Bisphenol A induces reactive oxygen species generation in the liver of male rats. Toxicology *188*, 117–124.

Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M.A., Jun, Z., Walter, J., and Jaenisch, R. (2002). Dnmt1 Overexpression Causes Genomic Hypermethylation, Loss of Imprinting, and Embryonic Lethality. Mol. Cell. Biol. *22*, 2124–2135.

Blattler, A., and Farnham, P.J. (2013). Cross-talk between Site-specific Transcription Factors and DNA Methylation States. Journal of Biological Chemistry *288*, 34287–34294.

Blewitt, M.E., Vickaryous, N.K., Paldi, A., Koseki, H., and Whitelaw, E. (2006). Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. PLoS Genet. *2*, e49.

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., and Lengauer, T. (2005). Bioinformatics *21*, 4067.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic - a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bonifati, V., Rizzu, P., van Baren, M.J., Schaap, O., Breedveld, G.J., Krieger, E., Dekker, M.C.J., Squitieri, F., Ibanez, P., Joosse, M., et al. (2003). Mutations in the DJ-1 gene associated with autosomal recessive early-onset parkinsonism. Science *299*, 256–259.

Borgel, J., Guibert, S., Li, Y., Chiba, H., Schübeler, D., Sasaki, H., Forné, T., and Weber, M. (2010). Targets and dynamics of promoter DNA methylation during early mouse development. Nat. Genet. *42*, 1093–1100.

Branco, M.R., Oda, M., and Reik, W. (2008). Safeguarding parental identity: Dnmt1 maintains imprints during epigenetic reprogramming in early embryogenesis. Genes Dev. *22*, 1567–1571.

Brennecke, J., Aravin, A.A., Stark, A., Dus, M., Kellis, M., Sachidanandam, R., and Hannon, G.J. (2007). Discrete Small RNA-Generating Loci as Master Regulators of Transposon Activity inDrosophila. Cell *128*, 1089–1103.

Brink, R.A. (1956). A Genetic Change Associated with the R Locus in Maize Which Is Directed and Potentially Reversible. Genetics *41*, 872–889.

Bromer, J.G., Zhou, Y., Taylor, M.B., Doherty, L., and Taylor, H.S. (2010). Bisphenol-A exposure in utero leads to epigenetic alterations in the developmental programming of uterine estrogen response. The FASEB Journal *24*, 2273–2280.

Bromfield, J.J. (2014). Seminal fluid and reproduction: much more than previously thought. J Assist Reprod Genet *31*, 627–636.

Bromfield, J.J., Schjenken, J.E., Chin, P.Y., Care, A.S., Jasper, M.J., and Robertson, S.A. (2014). Maternal tract factors contribute to paternal seminal fluid impact on metabolic phenotype in offspring. Proc Natl Acad Sci U S A *111*, 2200–2205.

Bultman, S.J., Michaud, E.J., and Woychik, R.P. (1992). Molecular characterization of the mouse agouti locus. Cell *71*, 1195–1204.

Burdge, G.C., Lillycrop, K.A., and Jackson, A.A. (2008). Nutrition in early life, and risk of cancer and metabolic disease: alternative endings in an epigenetic tale? Bjn *101*, 619–12.

Bygren, L.O., Tinghög, P., Carstensen, J., Edvinsson, S., Kaati, G., Pembrey, M.E., and Sjöström, M. (2014). Change in paternal grandmothers' early food supply influenced cardiovascular mortality of the female grandchildren. BMC Genet. *15*, 12.

Carone, B.R., Fauquier, L., Habib, N., Shea, J.M., Hart, C.E., Li, R., Bock, C., Li, C., Gu, H., Zamore, P.D., et al. (2010). Paternally Induced Transgenerational Environmental Reprogramming of Metabolic Gene Expression in Mammals. Cell *143*, 1084–1096.

Castel, S.E., and Martienssen, R.A. (2013). RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nat Rev Genet *14*, 100–112.

Chamorro-García, R., Sahu, M., Abbey, R.J., Laude, J., Pham, N., and Blumberg, B. (2013). Transgenerational inheritance of increased fat depot size, stem cell reprogramming, and hepatic steatosis elicited by prenatal exposure to the obesogen tributyltin in mice. Environ Health Perspect *121*, 359–366.

Chen, W., Goff, M.R., Kuang, H., and Chen, G. (2015). Higher Protein Kinase C ζ in Fatty Rat Liver and Its Effect on Insulin Actions in Primary Hepatocytes. PLoS ONE *10*, e0121890–17.

Chen, X., Weaver, J., Bove, B.A., Vanderveer, L.A., Weil, S.C., Miron, A., Daly, M.B., and Godwin, A.K. (2008). Allelic imbalance in BRCA1 and BRCA2 gene expression is associated with an increased breast cancer risk. Human Molecular Genetics *17*, 1336–1348.

Choi, S.-W., and Friso, S. (2010). Epigenetics: A New Bridge between Nutrition and Health. Advances in Nutrition: an International Review Journal *1*, 8–16.

Choi, S.H., Heo, K., Byun, H.-M., An, W., Lu, W., and Yang, A.S. (2010). Identification of preferential target sites for human DNA methyltransferases. Nucleic Acids Res. *39*, 104–118.

Chou, W.-C., Chen, J.-L., Lin, C.-F., Chen, Y.-C., Shih, F.-C., and Chuang, C.-Y. (2011). Biomonitoring of bisphenol A concentrations in maternal and umbilical cord blood in regard to birth outcomes and adipokine expression: a birth cohort study in Taiwan. Environ Health *10*, 94.

Chung, W.K., Shin, M., Jaramillo, T.C., Leibel, R.L., LeDuc, C.A., Fischer, S.G., Tzilianos, E., Gheith, A.A., Lewis, A.S., and Chetkovich, D.M. (2009). Absence epilepsy in apathetic, a spontaneous mutant mouse lacking the h channel subunit, HCN2. Neurobiol. Dis. *33*, 499–508.

Coe, E.H. (1959). A REGULAR AND CONTINUING CONVERSION-TYPE PHENOMENON AT THE B LOCUS IN MAIZE. Proc Natl Acad Sci U S A *45*, 828–832.

Court, F., Tayama, C., Romanelli, V., Martin-Trujillo, A., Iglesias-Platas, I., Okamura, K., Sugahara, N., Simon, C., Moore, H., Harness, J.V., et al. (2014). Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res. *24*, 554–569.

Cowley, M., and Oakey, R.J. (2012). Resetting for the next generation. Mol. Cell *48*, 819–821.

Crews, D., Gillette, R., Miller-Crews, I., Gore, A.C., and Skinner, M.K. (2014). Nature, nurture and epigenetics. Molecular and Cellular Endocrinology.

Crews, D., Gillette, R., Scarpino, S.V., Manikkam, M., Savenkova, M.I., and Skinner, M.K. (2012). Epigenetic transgenerational inheritance of altered stress responses. Proc Natl Acad Sci U S A *109*, 9143–9148.

Crews, D., Gore, A.C., Hsu, T.S., Dangleben, N.L., Spinetta, M., Schallert, T., Anway, M.D., and Skinner, M.K. (2007). Transgenerational epigenetic imprints on mate preference. Proc Natl Acad Sci U S A *104*, 5942–5946.

Cropley, J.E., Suter, C.M., Beckman, K.B., and Martin, D.I.K. (2006). Germ-line epigenetic modification of the murine A vy allele by nutritional supplementation. Proc Natl Acad Sci U S A *103*, 17308–17312.

Da Young Oh, Talukdar, S., Bae, E.J., Imamura, T., Morinaga, H., Fan, W., Li, P., Lu, W.J., Watkins, S.M., and Olefsky, J.M. (2010). GPR120 Is an Omega-3 Fatty Acid Receptor Mediating Potent Anti-inflammatory and Insulin-Sensitizing Effects. Cell *142*, 687–698.

Dai, H., and Wang, Z. (2014). Histone Modification Patterns and Their Responses to Environment. Curr Envir Health Rpt *1*, 11–21.

Dawson, T.M., Ko, H.S., and Dawson, V.L. (2010). Genetic Animal Models of Parkinson•s Disease. Neuron *66*, 646–661.

Daxinger, L., and Whitelaw, E. (2012). Understanding transgenerational epigenetic inheritance via the gametes in mammals. Nat Rev Genet.

de Vanssay, A., Bougé, A.-L., Boivin, A., Hermant, C., Teysset, L., Delmarre, V., Antoniewski, C., and Ronsseray, S. (2012). Paramutation in Drosophila linked to emergence of a piRNA-producing locus. Nature *490*, 112–115.

Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science *343*, 193–196.

Dhimolea, E., Wadia, P.R., Murray, T.J., Settles, M.L., Treitman, J.D., Sonnenschein, C., Shioda, T., and Soto, A.M. (2014). Prenatal Exposure to BPA Alters the Epigenome of the Rat Mammary Gland and Increases the Propensity to Neoplastic Development. PLoS ONE *9*, e99800–e99809.

Dickies, M.M. (1962). A new viable yellow mutation in the house mouse. J. Hered. *53*, 84–86.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dolinoy, D.C. (2008). The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. Nutr. Rev. *66 Suppl 1*, S7–S11.

Dolinoy, D.C., Das, R., Weidman, J.R., and Jirtle, R.L. (2007a). Metastable epialleles, imprinting, and the fetal origins of adult diseases. Pediatr. Res. *61*, 30R–37R.

Dolinoy, D.C., Huang, D., and Jirtle, R.L. (2007b). Maternal nutrient supplementation counteracts bisphenol A-induced DNA hypomethylation in early development. Proc Natl Acad Sci U S A *104*, 13056–13061.

Dolinoy, D.C., Weidman, J.R., Waterland, R.A., and Jirtle, R.L. (2006). Maternal Genistein Alters Coat Color and Protects Avy Mouse Offspring from Obesity by Modifying the Fetal Epigenome. Environ Health Perspect *114*, 567–572.

Doyle, T.J., Bowman, J.L., Windell, V.L., McLean, D.J., and Kim, K.H. (2013). Transgenerational effects of di-(2-ethylhexyl) phthalate on testicular germ cell associations and spermatogonial stem cells in mice. Biol. Reprod. *88*, 112–112.

Duhl, D.M.J., Vrieling, H., Miller, K.A., Wolff, G.L., and Barsh, G.S. (1994). Neomorphic agouti mutations in obese yellow mice. Nat. Genet. *8*, 59–65.

Eckersley-Maslin, M.A., Thybert, D., Bergmann, J.H., Marioni, J.C., Flicek, P., and Spector, D.L. (2014). Random Monoallelic Gene Expression Increases upon Embryonic Stem Cell Differentiation. Developmental Cell *28*, 351.

Edwards, A.W.F. (1963). The Measure of Association in a 2 × 2 Table. Journal of the Royal Statistical Society. Series a (General) *126*, 109.

Efron, B. (2007). Correlation and Large-Scale Simultaneous Significance Testing. Journal of the American Statistical Association *102*, 93–103.

Ekram, M.B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. Epigenetics *7*, 370–382.

Ellinghaus, D., Folseraas, T., Holm, K., Ellinghaus, E., Melum, E., Balschun, T., Laerdahl, J.K., Shiryaev, A., Gotthardt, D.N., Weismüller, T.J., et al. (2013). Genome-wide association analysis in Primary sclerosing cholangitis and ulcerative colitis identifies risk loci at GPR35and TCF4. Hepatology *58*, 1074–1083.

Emery, E.C., Young, G.T., Berrocoso, E.M., Chen, L., and McNaughton, P.A. (2011). HCN2 Ion Channels Play a Central Role in Inflammatory and Neuropathic Pain. Science *333*, 1462–.

Entringer, S., Epel, E.S., Kumsta, R., and Lin, J. (2011). Stress exposure in intrauterine life is associated with shorter telomere length in young adulthood.

Fang, F., Hodges, E., Molaro, A., Dean, M., Hannon, G.J., and Smith, A.D. (2012). Genomic landscape of human allele-specific DNA methylation. Proceedings of the National Academy of Sciences of the United States of America *109*, 7332–7337.

Faulk, C., Barks, A., and Dolinoy, D.C. (2013). Phylogenetic and DNA methylation analysis reveal novel regions of variable methylation in the mouse IAP class of transposons. BMC Genomics *14*, 1–1.

Faulk, C., Liu, K., Barks, A., Goodrich, J.M., and Dolinoy, D.C. (2014). Longitudinal epigenetic drift in mice perinatally exposed to lead. Epigenetics *9*, 934–941.

Faulkner, G.J., Kimura, Y., Daub, C.O., Wani, S., Plessy, C., Irvine, K.M., Schroder, K., Cloonan, N., Steptoe, A.L., Lassmann, T., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. Nat. Genet. *41*, 563–571.

Favorov, A., Mularoni, L., Cope, L.M., Medvedeva, Y., Mironov, A.A., Makeev, V.J., and Wheelan, S.J. (2012). Exploring Massive, Genome Scale Datasets with the GenometriCorr Package. PLoS Comput. Biol. *8*, e1002529–12.

Fay, M.P. (2010). Confidence intervals that match Fisher"s exact or Blaker"s exact tests. Biostatistics *11*, 373–374.

Feil, R., and Fraga, M.F. (2012). Epigenetics and the environment: emerging patterns and implications. Nat Rev Genet *13*, 97–109.

Fenton, S.E., Reed, C., and Newbold, R.R. (2012). Perinatal Environmental Exposures Affect Mammary Development, Function, and Cancer Risk in Adulthood. Annu. Rev. Pharmacol. Toxicol. *52*, 455–479.

Fernandez-Twinn, D.S., and Ozanne, S.E. (2006). Mechanisms by which poor early growth programs type-2 diabetes, obesity and the metabolic syndrome. Physiol. Behav. *88*, 234–243.

Ferreira, L.L., Couto, R., and Oliveira, P.J. (2014). Bisphenol A as epigenetic modulator: setting the stage for carcinogenesis? Eur J Clin Invest *45*, 32–36.

Forrest, M.P., Hill, M.J., Quantock, A.J., Martin-Rendon, E., and Blake, D.J. (2014). The emerging roles of TCF4 in disease and development. Trends in Molecular Medicine *20*, 322–331.

Fullston, T., Ohlsson Teague, E.M.C., Palmer, N.O., DeBlasio, M.J., Mitchell, M., Corbett, M., Print, C.G., Owens, J.A., and Lane, M. (2013). Paternal obesity initiates metabolic disturbances in two generations of mice with incomplete penetrance to the F2 generation and alters the transcriptional profile of testis and sperm microRNA content. Faseb J. *27*, 4226–4243.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Gaudet, F., Rideout, W.M., Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004). Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. Mol. Cell. Biol. *24*, 1640–1648.

Gaujoux, R., and Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. BMC Bioinformatics *11*, 367.

Gendrel, A.-V., Attia, M., Chen, C.-J., Diabangouaya, P., Servant, N., Barillot, E., and Heard, E. (2014). Developmental Dynamics and Disease Potential of Random Monoallelic Gene Expression. Developmental Cell *28*, 366–380.

Gendrel, A.-V., Marion-Poll, L., Katoh, K., and Heard, E. (2016). Random monoallelic expression of genes on autosomes: Parallels with X-chromosome inactivation. Seminars in Cell and Developmental Biology *56*, 100–110.

Gheni, G., Ogura, M., Iwasaki, M., Yokoi, N., Minami, K., Nakayama, Y., Harada, K., Hastoy, B., Wu, X., Takahashi, H., et al. (2014). Glutamate Acts as a Key Signal Linking Glucose Metabolism to Incretin/cAMP Action to Amplify Insulin Secretion. Cell Rep *9*, 661–673.

Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells. Cell *153*, 1149–1163.

Gillette, R., Miller-Crews, I., Nilsson, E.E., Skinner, M.K., Gore, A.C., and Crews, D. (2014). Sexually dimorphic effects of ancestral exposure to vinclozolin on stress reactivity in rats. Endocrinology *155*,

3853–3866.

Gluckman, P.D., and Hanson, M.A. (2004). The developmental origins of the metabolic syndrome. Trends Endocrinol. Metab. *15*, 183–187.

Grandjean, V., Gounon, P., Wagner, N., Martin, L., Wagner, K.D., Bernex, F., Cuzin, F., and Rassoulzadegan, M. (2009). The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. Development *136*, 3647–3655.

Gruenbaum, Y., Cedar, H., and Razin, A. (1982). Substrate and sequence specificity of a eukaryotic DNA methylase. Nature *295*, 620–622.

Guerrero-Bosagna, C., Covert, T.R., Haque, M.M., Settles, M., Nilsson, E.E., Anway, M.D., and Skinner, M.K. (2012). Epigenetic transgenerational inheritance of vinclozolin induced mouse adult onset disease and associated sperm epigenome biomarkers. Reproductive Toxicology *34*, 694–707.

Guerrero-Bosagna, C., Savenkova, M., Haque, M.M., Nilsson, E., and Skinner, M.K. (2013). Environmentally induced epigenetic transgenerational inheritance of altered Sertoli cell transcriptome and epigenome: molecular etiology of male infertility. PLoS ONE *8*, e59922.

Guerrero-Bosagna, C., Settles, M., Lucker, B., and Skinner, M.K. (2010). Epigenetic transgenerational actions of vinclozolin on promoter regions of the sperm epigenome. PLoS ONE *5*, e13100.

Guibert, S., Forné, T., and Weber, M. (2012). Global profiling of DNA methylation erasure in mouse primordial germ cells. Genome Res. *22*, 633–641.

Hackett, J.A., and Surani, M.A. (2012). DNA methylation dynamics during the mammalian life cycle. Philosophical Transactions of the Royal Society B: Biological Sciences *368*, 20110328–20110328.

Hackett, J.A., Dietmann, S., Murakami, K., Down, T.A., Leitch, H.G., and Surani, M.A. (2013a). Synergistic Mechanisms of DNA Demethylation during Transition to Ground-State Pluripotency. Stem Cell Reports *1*, 518–531.

Hackett, J.A., Sengupta, R., Zylicz, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013b). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science *339*, 448–452.

Hales, C.N., and Barker, D.J. (2001). The thrifty phenotype hypothesis. Br. Med. Bull. *60*, 5–20.

Hang, H., Zhang, Y., Dunbrack, R.L., Jr., Wang, C., and Lieberman, H.B. (2002). Identification and Characterization of a Paralog of Human Cell Cycle Checkpoint Gene HUS1. Genomics *79*, 487–492.

Hansen, K.D., Langmead, B., and Irizarry, R.A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. Genome Biol *13*, R83.

Hansen, K.D., Sabunciyan, S., Langmead, B., Nagy, N., Curley, R., Klein, G., Klein, E., Salamon, D., and Feinberg, A.P. (2014). Large-scale hypomethylated blocks associated with Epstein-Barr virus-induced B-cell immortalization. Genome Res. *24*, 177–184.

Harley, K.G., Aguilar Schall, R., Chevrier, J., Tyler, K., Aguirre, H., Bradman, A., Holland, N.T., Lustig, R.H., Calafat, A.M., and Eskenazi, B. (2013). Prenatal and Postnatal Bisphenol A Exposure and Body Mass Index in Childhood in the CHAMACOS Cohort. Environ Health Perspect 1–7.

Harten, S.K. (2015). The recently identified modifier of murine metastable epialleles, Rearranged L-Myc Fusion, is involved in maintaining epigenetic marks at CpG island shores and enhancers. BMC Biol *13*, 1–

15.

Hatch, E.E., Nelson, J.W., Stahlhut, R.W., and Webster, T.F. (2010). Association of endocrine disruptors and obesity: perspectives from epidemiological studies. International Journal of Andrology *33*, 324–332.

Heard, E., and Martienssen, R.A. (2014). Transgenerational epigenetic inheritance: myths and mechanisms. Cell *157*, 95–109.

Heijmans, B.T., Tobi, E.W., Stein, A.D., Putter, H., Blauw, G.J., Susser, E.S., Eline Slagboom, P., and Lumey, L.H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. pp. 17046–17049.

Hernandez, A.I., Blace, N., Crary, J.F., Serrano, P.A., Leitges, M., Libien, J.M., Weinstein, G., Tcherapanov, A., and Sacktor, T.C. (2011). Protein Kinase M Synthesis from a Brain mRNA Encoding an Independent Protein Kinase C Catalytic Domain: IMPLICATIONS FOR THE MOLECULAR MECHANISM OF MEMORY. Journal of Biological Chemistry *278*, 40305–40316.

Heydemann, A. (2016). An Overview of Murine High Fat Diet as a Model for Type 2 Diabetes Mellitus. Journal of Diabetes Research *2016*, 2902351–14.

Hilakivi-Clarke, L. (2014). Maternal exposure to diethylstilbestrol during pregnancy and increased breast cancer risk in daughters. Breast Cancer Research (Online Edition) *16*, 208–208.

Hirasawa, R., Chiba, H., Kaneda, M., Tajima, S., Li, E., Jaenisch, R., and Sasaki, H. (2008). Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. Genes Dev. *22*, 1607–1616.

Ho, S.M. (2006). Developmental Exposure to Estradiol and Bisphenol A Increases Susceptibility to Prostate Carcinogenesis and Epigenetically Regulates Phosphodiesterase Type 4 Variant 4. Cancer Research *66*, 5624–5632.

Ho, S.-M., Cheong, A., Adgent, M.A., Veevers, J., Suen, A.A., Tam, N.N.C., Leung, Y.-K., Jefferson, W.N., and Williams, C.J. (2017). Environmental factors, epigenetics, and developmental origin of reproductive disorders. Reproductive Toxicology *68*, 85–104.

Ichimura, A., Hirasawa, A., Poulain-Godefroy, O., Bonnefond, A., Hara, T., Yengo, L., Kimura, I., Leloire, A., Liu, N., Iida, K., et al. (2012). Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. Nature *483*, 350–357.

Inawaka, K., Kawabe, M., Takahashi, S., Doi, Y., Tomigahara, Y., Tarui, H., Abe, J., Kawamura, S., and Shirai, T. (2009). Maternal exposure to anti-androgenic compounds, vinclozolin, flutamide and procymidone, has no effects on spermatogenesis and DNA methylation in male rats of subsequent generations. Toxicology and Applied Pharmacology *237*, 178–187.

Iqbal, K., Tran, D.A., Li, A.X., Warden, C., Bai, A.Y., Singh, P., Wu, X., Pfeifer, G.P., and Szabó, P.E. (2015). Deleterious effects of endocrine disruptors are corrected in the mammalian germline by epigenome reprogramming. Genome Biol *16*, 59.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Ishiuchi, T., and Torres-Padilla, M.-E. (2014). LINEing germ and embryonic stem cells' silencing of retrotransposons. Genes Dev. *28*, 1381–1383.

Jaffe, A.E., Murakami, P., Lee, H., Leek, J.T., Fallin, M.D., Feinberg, A.P., and Irizarry, R.A. (2012). Bump hunting to identify differentially methylated regions in epigenetic epidemiology studies. Int J Epidemiol *41*, 200–209.

Jeffries, M.A., and Sawalha, A.H. (2011). Epigenetics in systemic lupus erythematosus: leading the way for specific therapeutic agents. Int J Clin Rheumtol *6*, 423–439.

Jenkins, S., Raghuraman, N., Eltoum, I., Carpenter, M., Russo, J., and Lamartiniere, C.A. (2009). Oral Exposure to Bisphenol A Increases Dimethylbenzanthracene-Induced Mammary Cancer in Rats. Environ Health Perspect *117*, 910–915.

Jirtle, R.L., and Skinner, M.K. (2007). Environmental epigenomics and disease susceptibility. Nat Rev Genet *8*, 253–262.

Kang, E.-R., Iqbal, K., Tran, D.A., Rivas, G.E., Singh, P., Pfeifer, G.P., and Szab, P.E. (2014). Effects of endocrine disruptors on imprinted gene expression in the mouse embryo. Epigenetics *6*, 937–950.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. Nucleic Acids Res. *32*, D493–D496.

Kawashima, T., and Berger, F. (2014). Epigenetic reprogramming in plant sexual reproduction. Nat Rev Genet *15*, 613–624.

Ke, Z.-H., Pan, J.-X., Jin, L.-Y., Xu, H.-Y., Yu, T.-T., Ullah, K., Rahman, T.U., Ren, J., Cheng, Y., Dong, X.-Y., et al. (2016). Bisphenol A Exposure May Induce Hepatic Lipid Accumulation via Reprogramming the DNA Methylation Patterns of Genes Involved in Lipid Metabolism. Nature Publishing Group 1–13.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res. *12*, 996–1006.

Khurana, J.S., Wang, J., Xu, J., Koppetsch, B.S., Thomson, T.C., Nowosielska, A., Li, C., Zamore, P.D., Weng, Z., and Theurkauf, W.E. (2011). Adaptation to P element transposon invasion in Drosophila melanogaster. Cell *147*, 1551–1563.

Kiani, J., Grandjean, V., Liebers, R., Tuorto, F., Ghanbarian, H., Lyko, F., Cuzin, F., and Rassoulzadegan, M. (2013). RNA–mediated epigenetic heredity requires the cytosine methyltransferase Dnmt2. PLoS Genet. *9*, e1003498.

Kim, S.-J., Miller, J.L., Kuipers, P.J., German, J.R., Beaudet, A.L., Sahoo, T., and Driscoll, D.J. (2012). Unique and atypical deletions in Prader-Willi syndrome reveal distinct phenotypes. Eur. J. Hum. Genet. *20*, 283–290.

Klip, H., Verloop, J., van Gool, J.D., Koster, M.E., Burger, C.W., and van Leeuwen, F.E. (2002). Hypospadias in sons of women exposed to diethylstilbestrol in utero: a cohort study. The Lancet *359*, 1102–1107.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Laing, L.V., Viana, J., Dempster, E.L., Trznadel, M., Trunkfield, L.A., Webster, T.M.U., van Aerle, R., Paull, G.C., Wilson, R.J., Mill, J., et al. (2016). Bisphenol A causes reproductive toxicity, decreases dnmt1

transcription, and reduces global DNA methylation in breeding zebrafish \(Danio rerio\). Epigenetics *11*, 526–538.

Lambertini, L. (2014). Genomic imprinting. Current Opinion in Pediatrics *26*, 237–242.

Lang, I.A., Galloway, T.S., Scarlett, A., Henley, W.E., Depledge, M., Wallace, R.B., and Melzer, D. (2008). Association of urinary bisphenol A concentration with medical disorders and laboratory abnormalities in adults. Jama *300*, 1303–1310.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth *9*, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lay, F.D., Liu, Y., Kelly, T.K., Witt, H., Farnham, P.J., Jones, P.A., and Berman, B.P. (2015). The role of DNA methylation in directing the functional organization of the cancer epigenome. Genome Res. *25*, 467–477.

Leonhardt, H., Page, A.W., Weier, H.U., and Bestor, T.H. (1992). A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. Cell *71*, 865–873.

Leung, A., Trac, C., Du, J., Natarajan, R., and Schones, D.E. (2016). Persistent Chromatin Modifications Induced by High Fat Diet. J. Biol. Chem. *291*, 10446–10455.

Levin, H.L., and Moran, J.V. (2011). Dynamic interactions between transposable elements and their hosts. Nat Rev Genet *12*, 627.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, G., Chang, H., Xia, W., Mao, Z., Li, Y., and Xu, S. (2014). F0 maternal BPA exposure induced glucose intolerance of F2 generation through DNA methylation change in Gck. Toxicology Letters *228*, 192–199.

Li, L., Lu, X., and Dean, J. (2013). The maternal to zygotic transition in mammals. Molecular Aspects of Medicine *34*, 919–938.

Li, M., Cai, S.-Y., and Boyer, J.L. (2017). Mechanisms of bile acid mediated inflammation in the liver. Molecular Aspects of Medicine 1–9.

Li, Z., Dai, H., Martos, S.N., Xu, B., Gao, Y., Li, T., Zhu, G., Schones, D.E., and Wang, Z. (2015). Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. Genome Biol *16*, 115–130.

Liang, G., Chan, M.F., Tomigahara, Y., Tsai, Y.C., Gonzales, F.A., Li, E., Laird, P.W., and Jones, P.A. (2002). Cooperativity between DNA Methyltransferases in the Maintenance Methylation of Repetitive Elements. Mol. Cell. Biol. *22*, 480–491.

Lo, H.S., Wang, Z., Hu, Y., Yang, H.H., Gere, S., Buetow, K.H., and Lee, M.P. (2003). Allelic variation in gene expression is common in the human genome. Genome Res. *13*, 1855–1862.

Lorincz, M.C., Dickerson, D.R., Schmitt, M., and Groudine, M. (2004). Intragenic DNA methylation alters

chromatin structure and elongation efficiency in mammalian cells. Nature Structural & Molecular Biology *11*, 1068–1075.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550–21.

Ludwig, A., Budde, T., Stieber, J., Moosmang, S., Wahl, C., Holthoff, K., Langebartels, A., Wotjak, C., Munsch, T., Zong, X., et al. (2003). Absence epilepsy and sinus dysrhythmia in mice lacking the pacemaker channel HCN2. Embo J *22*, 216–224.

MacDonald, W.A., and Mann, M.R.W. (2013). Epigenetic regulation of genomic imprinting from germ line to preimplantation. Mol. Reprod. Dev. *81*, 126–140.

Maffini, M.V., Rubin, B.S., Sonnenschein, C., and Soto, A.M. (2006). Endocrine disruptors and reproductive health: The case of bisphenol-A. Molecular and Cellular Endocrinology *254-255*, 179–186.

Manikkam, M., Haque, M.M., and Guerrero-Bosagna, C. (2014). Pesticide Methoxychlor Promotes the Epigenetic Transgenerational Inheritance of Adult-Onset Disease through the Female Germline. PLoS ONE.

Manikkam, M., Guerrero-Bosagna, C., Tracey, R., Haque, M.M., and Skinner, M.K. (2012a). Transgenerational Actions of Environmental Compounds on Reproductive Disease and Identification of Epigenetic Biomarkers of Ancestral Exposures. PLoS ONE *7*, e31901.

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M.K. (2012b). Dioxin (TCDD) induces epigenetic transgenerational inheritance of adult onset disease and sperm epimutations. PLoS ONE *7*, e46249.

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M.K. (2012c). Pesticide and insect repellent mixture (permethrin and DEET) induces epigenetic transgenerational inheritance of disease and sperm epimutations. Reprod. Toxicol. *34*, 708–719.

Manikkam, M., Tracey, R., Guerrero-Bosagna, C., and Skinner, M.K. (2013). Plastics derived endocrine disruptors (BPA, DEHP and DBP) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. PLoS ONE *8*, e55387.

Martin, I., Kim, J.W., Lee, B.D., Kang, H.C., Xu, J.-C., Jia, H., Stankowski, J., Kim, M.-S., Zhong, J., Kumar, M., et al. (2014). Ribosomal Protein s15 Phosphorylation Mediates LRRK2 Neurodegeneration in Parkinson's Disease. Cell *157*, 472–485.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 10.

Martos, S.N., Tang, W.-Y., and Wang, Z. (2015). Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. Prog. Biophys. Mol. Biol. *118*, 44–54.

Maunakea, A.K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. Nature Publishing Group *23*, 1256–1269.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.

McCarrey, J.R. (2012). The epigenome as a target for heritable environmental disruptions of cellular

function. Molecular and Cellular Endocrinology *354*, 9–15.

McCarrey, J.R. (2014). Distinctions between transgenerational and non-transgenerational epimutations. Molecular and Cellular Endocrinology.

McClellan, J., and King, M.-C. (2010). Genetic Heterogeneity in Human Disease. Cell *141*, 210–217.

McLachlan, J.A. (2016). Environmental signaling: from environmental estrogens to endocrine-disrupting chemicals and beyond. Andrology *4*, 684–694.

McRae, A.F., Powell, J.E., Henders, A.K., Bowdler, L., Hemani, G., Shah, S., Painter, J.N., Martin, N.G., Visscher, P.M., and Montgomery, G.W. (2014). Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol *15*, R73.

Melzer, D., Rice, N.E., Lewis, C., Henley, W.E., and Galloway, T.S. (2010). Association of Urinary Bisphenol A Concentration with Heart Disease: Evidence from NHANES 2003/06. PLoS ONE *5*, e8673–e8679.

Michaud, E.J., van Vugt, M.J., Bultman, S.J., Sweet, H.O., Davisson, M.T., and Woychik, R.P. (1994). Differential expression of a new dominant agouti allele (Aiapy) is correlated with methylation state and is influenced by parental lineage. Genes Dev. *8*, 1463–1472.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Miyawaki, J., Sakayama, K., Kato, H., Yamamoto, H., and Masuno, H. (2007). Perinatal and postnatal exposure to bisphenol a increases adipose tissue mass and serum cholesterol level in mice. J. Atheroscler. Thromb. *14*, 245–252.

Morgan, H.D.H., Sutherland, H.G.H., Martin, D.I.D., and Whitelaw, E.E. (1999). Epigenetic inheritance at the agouti locus in the mouse. Nat. Genet. *23*, 314–318.

Mosteller, F. (1968). Association and Estimation in Contingency Tables. Journal of the American Statistical Association *63*, 1–28.

Muerdter, F., Olovnikov, I., Molaro, A., Rozhkov, N.V., Czech, B., Gordon, A., Hannon, G.J., and Aravin, A.A. (2012). Production of artificial piRNAs in flies and mice. Rna *18*, 42–52.

Murphy, S.K., and Hoyo, C. (2013). Sculpting Our Future: Environmental Nudging of the Imprintome. In Environmental Epigenomics in Health and Disease, R.L. Jirtle, and F.L. Tyson, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 51–73.

Nakagami, A., Negishi, T., Kawasaki, K., Imai, N., Nishida, Y., Ihara, T., Kuroda, Y., Yoshikawa, Y., and Koyama, T. (2009). Alterations in male infant behaviors towards its mother by prenatal exposure to bisphenol A in cynomolgus monkeys (Macaca fascicularis) during early suckling period. Psychoneuroendocrinology *34*, 1189–1197.

NCBI Resource Coordinators (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. *44*, D7–D19.

Nichols, J., and Smith, A. (2009). Naive and Primed Pluripotent States. Stem Cell *4*, 487–492.

Nilsson, E.E., and Skinner, M.K. (2014). Environmentally induced epigenetic transgenerational inheritance of disease susceptibility. Transl Res *0*.

Nilsson, E.E., Anway, M.D., Stanfield, J., and Skinner, M.K. (2008). Transgenerational epigenetic effects of the endocrine disruptor vinclozolin on pregnancies and female adult onset disease. Reproduction *135*, 713–721.

Nilsson, E., Larsen, G., Manikkam, M., Guerrero-Bosagna, C., Savenkova, M.I., and Skinner, M.K. (2012). Environmentally induced epigenetic transgenerational inheritance of ovarian disease. PLoS ONE *7*, e36129.

O'Hagan, H.M., and Tang, W.-Y. (2013). Increased understanding of the impact of environmental exposures on the epigenome. Environ. Mol. Mutagen. *55*, 151–154.

Ohta, T., Gray, T.A., Rogan, P.K., Buiting, K., Gabriel, J.M., Saitoh, S., Muralidhar, B., Bilienska, B., Krajewska-Walasek, M., Driscoll, D.J., et al. (1999). Imprinting-mutation mechanisms in Prader-Willi syndrome. Am. J. Hum. Genet. *64*, 397–413.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. Cell *99*, 247–257.

Padmanabhan, N., Jia, D., Geary-Joo, C., Wu, X., Ferguson-Smith, A.C., Fung, E., Bieda, M.C., Snyder, F.F., Gravel, R.A., Cross, J.C., et al. (2013). Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. Cell *155*, 81–93.

Pembrey, M., Saffery, R., Bygren, L.O., Network in Epigenetic Epidemiology (2014). Human transgenerational responses to early-life experience: potential impact on development, health and biomedical research. Journal of Medical Genetics *51*, 563–572.

Peng, Q., and Ecker, J.R. (2012). Detection of allele-specific methylation through a generalized heterogeneous epigenome model. Bioinformatics *28*, i163–i171.

Perera, F., and Herbstman, J. (2011). Prenatal environmental exposures, epigenetics, and disease. Reproductive Toxicology *31*, 363–373.

Piedrahita, J.A. (2011). The role of imprinted genes in fetal growth abnormalities. Birth Defects Research Part a: Clinical and Molecular Teratology *91*, 682–692.

Prins, G.S., and Ho, S.M. (2010). Early-life estrogens and prostate cancer in an animal model. J Devel Orig Health Dis *1*, 365–370.

Radford, E.J., Ito, M., Shi, H., Corish, J.A., Yamazawa, K., Isganaitis, E., Seisenberger, S., Hore, T.A., Reik, W., Erkek, S., et al. (2014). In utero undernourishment perturbs the adult sperm methylome and intergenerational metabolism. Science *345*, 1255903–1255903.

Rakyan, V.K., Blewitt, M.E., Druker, R., Preis, J.I., and Whitelaw, E. (2002). Metastable epialleles in mammals. Trends in Genetics *18*, 348–351.

Rakyan, V.K., PREIS, J., Morgan, H.D., and Whitelaw, E. (2001). The marks, mechanisms and memory of epigenetic states in mammals. Biochem. J. *356*, 1.

Rappaport, S.M., and Smith, M.T. (2010). Epidemiology. Environment and disease risks. Science *330*, 460–461.

Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature *441*, 469–474.

Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P., and Vinson, C.

(2010). CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc Natl Acad Sci U S A *107*, 20311–20316.

Robertson, S.A. (2005). Seminal plasma and male factor signalling in the female reproductive tract. Cell Tissue Res *322*, 43–52.

Roche, S.E., and Rio, D.C. (1998). Trans-silencing by P elements inserted in subtelomeric heterochromatin involves the Drosophila Polycomb group gene, Enhancer of zeste. Genetics *149*, 1839.

Ronsseray, S., Josse, T., Boivin, A., and Anxolabéhère, D. (2003). Telomeric transgenes and trans-silencing in Drosophila. Genetica *117*, 327–335.

Rugg-Gunn, P.J., Ferguson-Smith, A.C., and Pedersen, R.A. (2007). Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. Human Molecular Genetics *16*, R243–R251.

Saal, vom, F.S., and Hughes, C. (2005). An Extensive New Literature Concerning Low-Dose Effects of Bisphenol A Shows the Need for a New Risk Assessment. Environ Health Perspect *113*, 926–933.

Saxena, R., Hivert, M.-F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U., et al. (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat. Genet. *42*, 142–148.

Schmidt, C.W. (2013). Uncertain Inheritance: Transgenerational Effects of Environmental Exposures. Environ Health Perspect *121*, A298–A303.

SCHNEIDER, S., KAUFMANN, W., BUESEN, R., and VANRAVENZWAAY, B. (2008). Vinclozolin—The lack of a transgenerational effect after oral maternal exposure during organogenesis. Reproductive Toxicology *25*, 352–360.

Schneider, S., Marxfeld, H., Gröters, S., Buesen, R., and van Ravenzwaay, B. (2013). Vinclozolin—No transgenerational inheritance of anti-androgenic effects after maternal exposure during organogenesis via the intraperitoneal route. Reproductive Toxicology *37*, 6–14.

Schübeler, D. (2012). Molecular biology. Epigenetic islands in a genetic ocean. Science *338*, 756–757.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell *48*, 849–862.

Seisenberger, S., Popp, C., and Reik, W. (2010). Retrotransposons and germ cells: reproduction, death, and diversity. F1000 Biol Rep *2*.

Selvaraj, A., Balamurugan, K., Yepiskoposyan, H., Zhou, H., Egli, D., Georgiev, O., Thiele, D.J., and Schaffner, W. (2005). Metal-responsive transcription factor (MTF-1) handles both extremes, copper load and copper starvation, by activating different genes. Genes Dev. *19*, 891–896.

Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79.

Simón-Sánchez, J., and Singleton, A. (2008). Genome-wide association studies in neurological disorders. The Lancet Neurology *7*, 1067–1072.

Sims, H.I., Chirn, G.W., and Marr, M.T. (2012). Single nucleotide in the MTF-1 binding site can determine

metal-specific transcription activation.

Singh, S., and Li, S.S.-L. (2012). Epigenetic effects of environmental chemicals bisphenol A and phthalates. Ijms *13*, 10143–10153.

Skaar, D.A., Li, Y., Bernal, A.J., Hoyo, C., Murphy, S.K., and Jirtle, R.L. (2012). The human imprintome: regulatory mechanisms, methods of ascertainment, and roles in disease susceptibility. Ilar J *53*, 341–358.

Skinner, M.K. (2008). What is an epigenetic transgenerational phenotype? Reproductive Toxicology *25*, 2–6.

Skinner, M.K. (2013). Environmental Epigenetics and Epigenetic Transgenerational Inheritance. In Environmental Epigenomics in Health and Disease, (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 245–256.

Skinner, M.K. (2014). Endocrine disruptor induction of epigenetic transgenerational inheritance of disease. Molecular and Cellular Endocrinology 1–9.

Skinner, M.K., Anway, M.D., Savenkova, M.I., Gore, A.C., and Crews, D. (2008). Transgenerational Epigenetic Programming of the Brain Transcriptome and Anxiety Behavior. PLoS ONE *3*, e3745.

Skinner, M.K., Haque, C.G.-B.M., Nilsson, E., Bhandari, R., and McCarrey, J.R. (2013a). Environmentally induced transgenerational epigenetic reprogramming of primordial germ cells and the subsequent germ line. PLoS ONE *8*, e66318.

Skinner, M.K., Manikkam, M., Tracey, R., Guerrero-Bosagna, C., Haque, M., and Nilsson, E.E. (2013b). Ancestral dichlorodiphenyltrichloroethane (DDT) exposure promotes epigenetic transgenerational inheritance of obesity. BMC Med *11*, 228.

Skinner, M.K., Savenkova, M.I., Zhang, B., Gore, A.C., and Crews, D. (2014). Gene bionetworks involved in the epigenetic transgenerational inheritance of altered mate preference: environmental epigenetics and evolutionary biology. BMC Genomics *15*, 377.

Skryabin, B.V., Gubar, L.V., Seeger, B., Pfeiffer, J., Handel, S., Robeck, T., Karpova, E., Rozhdestvensky, T.S., and Brosius, J. (2007). Deletion of the MBII-85 snoRNA gene cluster in mice results in postnatal growth retardation. PLoS Genet. *3*, e235.

Smallwood, S.A., Tomizawa, S.-I., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat. Genet. *43*, 811–814.

Smit, A., Hubley, R., and Green, P. (2013). RepeatMasker: Open-4.0. <http://www.repeatmasker.org>. 1–2.

Smith, F.M., Garfield, A.S., and Ward, A. (2006). Regulation of growth and metabolism by imprinted genes. Cytogenet Genome Res *113*, 279–291.

Smith, Z.D., Chan, M.M., Humm, K.C., Karnik, R., Mekhoubad, S., Regev, A., Eggan, K., and Meissner, A. (2014). DNA methylation dynamics of the human preimplantation embryo. Nature *511*, 611–615.

Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature *484*, 339–344.

Somm, E., Schwitzgebel, V.R.M., Toulotte, A., Cederroth, C.R., Combescure, C., Nef, S., Aubert, M.L., and H ppi, P.S. (2009). Perinatal Exposure to Bisphenol A Alters Early Adipogenesis in the Rat. Environ

Health Perspect *117*, 1549–1555.

Soto, A.M., Brisken, C., Schaeberle, C., and Sonnenschein, C. (2013). Does Cancer Start in the Womb? Altered Mammary Gland Development and Predisposition to Breast Cancer due to in Utero Exposure to Endocrine Disruptors. J Mammary Gland Biol Neoplasia *18*, 199–208.

Soubry, A., Hoyo, C., Jirtle, R.L., and Murphy, S.K. (2014). A paternal environmental legacy: evidence for epigenetic inheritance through the male germ line. BioEssays *36*, 359–371.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G., and Leonhardt, H. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. The Journal of Cell Biology *176*, 565–571.

Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. pp. 3418–3422.

Stelzer, Y., Ronen, D., Bock, C., Boyle, P., Meissner, A., and Benvenisty, N. (2013). Stem Cell Reports *1*, 79.

Stojanoska, M.M., Milosevic, N., Milic, N., and Abenavoli, L. (2017). The influence of phthalates and bisphenol A on the obesity development and glucose metabolism disorders. Endocrine 1–16.

Stouder, C., and Paoloni-Giacobino, A. (2010). Transgenerational effects of the endocrine disruptor vinclozolin on the methylation pattern of imprinted genes in the mouse sperm. Reproduction *139*, 373–379.

Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics *24*, 1461–1462.

Stuwe, E., Tóth, K.F., and Aravin, A.A. (2014). Small but sturdy: small RNAs in cellular memory and epigenetics. Genes Dev. *28*, 423–431.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545–15550.

Susiarjo, M., Sasson, I., Mesaros, C., and Bartolomei, M.S. (2013). Bisphenol A Exposure Disrupts Genomic Imprinting in the Mouse. PLoS Genet. *9*.

Susiarjo, M., Xin, F., Bansal, A., Stefaniak, M., Li, C., Simmons, R.A., and Bartolomei, M.S. (2015). Bisphenol A exposure disrupts metabolic health across multiple generations in the mouse. Endocrinology en.2014–2027–10.

Susiarjo, M., Xin, F., Stefaniak, M., Mesaros, C., Simmons, R.A., and Bartolomei, M.S. (2017). Bile acids and tryptophan metabolism are novel pathways involved in metabolic abnormalities in BPA-exposed pregnant mice and male offspring. Endocrinology 1–17.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033–3042.

Tang, W.-Y., and Ho, S.-M. (2007). Epigenetic reprogramming and imprinting in origins of disease. Rev Endocr Metab Disord *8*, 173–182.

Tang, W.-Y., Morey, L.M., Cheung, Y.Y., Birch, L., Prins, G.S., and Ho, S.-M. (2012). Neonatal Exposure to Estradiol/Bisphenol A Alters Promoter Methylation and Expression of Nsbp1and Hpcal1Genes and Transcriptional Programs of Dnmt3a/band Mbd2/4in the RatProstate Gland Throughout Life.

Endocrinology *153*, 42–55.

Taoka, H., Yokoyama, Y., Morimoto, K., Kitamura, N., Tanigaki, T., Takashina, Y., Tsubota, K., and Watanabe, M. (2016). Role of bile acids in the regulation of the metabolic pathways. World J Diabetes *7*, 260–270.

Thayer, K.A., Heindel, J.J., Bucher, J.R., and Gallo, M.A. (2012). Role of Environmental Chemicals in Diabetes and Obesity: A National Toxicology Program Workshop Review. Environ Health Perspect *120*, 779–789.

Thorvaldsen, J.L., Mann, M.R.W., Nwoko, O., Duran, K.L., and Bartolomei, M.S. (2002). Analysis of Sequence Upstream of the Endogenous H19 Gene Reveals Elements Both Essential and Dispensable for Imprinting. Mol. Cell. Biol. *22*, 2450–2462.

Tracey, R., Manikkam, M., Guerrero-Bosagna, C., and Skinner, M.K. (2013). Hydrocarbons (jet fuel JP-8) induce epigenetic transgenerational inheritance of obesity, reproductive disease and sperm epimutations. Reproductive Toxicology *36*, 104–116.

Trasler, J.M. (2006). Gamete imprinting: setting epigenetic patterns for the next generation. Reprod. Fertil. Dev. *18*, 63.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S.-I., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H.R., et al. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. Genes to Cells *11*, 805–814.

Tucker, K.L., Beard, C., Dausmann, J., Jackson-Grusby, L., Laird, P.W., Lei, H., Li, E., and Jaenisch, R. (1996). Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. Genes Dev. *10*, 1008–1020.

Tycko, B. (2010). Allele-specific DNA methylation: beyond imprinting. Human Molecular Genetics *19*, R210–R220.

van Esterik, J.C.J., Dollé, M.E.T., Lamoree, M.H., van Leeuwen, S.P.J., Hamers, T., Legler, J., and van der Ven, L.T.M. (2014). Programming of metabolic effects in C57BL/6JxFVB mice by exposure to bisphenol A during gestation and lactation. Toxicology *321*, 40–52.

van Esterik, J.C.J., Vitins, A.P., Hodemaekers, H.M., Kamstra, J.H., Legler, J., Pennings, J.L.A., Steegenga, W.T., Lute, C., Jelinek, J., Issa, J.P.J., et al. (2015). Liver DNA methylation analysis in adult female C57BL/6JxFVB mice following perinatal exposure to bisphenol A. Toxicology Letters *232*, 293–300.

Veenendaal, M.V.E., Painter, R.C., de Rooij, S.R., Bossuyt, P.M.M., van der Post, J.A.M., Gluckman, P.D., Hanson, M.A., and Roseboom, T.J. (2013). Transgenerational effects of prenatal exposure to the 1944-45 Dutch famine. Bjog *120*, 548–553.

Veenendaal, M.V., Costello, P.M., Lillycrop, K.A., de Rooij, S.R., van der Post, J.A., Bossuyt, P.M., Hanson, M.A., Painter, R.C., and Roseboom, T.J. (2012). Prenatal famine exposure, health in later life and promoter methylation of four candidate genes. J Devel Orig Health Dis *3*, 450–457.

Vigé, A., Gallou-Kabani, C., and Junien, C. (2008). Sexual Dimorphism in Non-Mendelian Inheritance. Pediatr. Res. *63*, 340–347.

Visscher, P.M., Brown, M.A., McCarthy, M.I., and Yang, J. (2012). REVIEW Five Years of GWAS Discovery. Am. J. Hum. Genet. *90*, 7–24.

Wagner, K.D., Wagner, N., Ghanbarian, H., Grandjean, V., Gounon, P., Cuzin, F., and Rassoulzadegan, M. (2008). RNA Induction and Inheritance of Epigenetic Cardiac Hypertrophy in the Mouse. Developmental Cell *14*, 962–969.

Walsh, C.P., Chaillet, J.R., and Bestor, T.H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat. Genet. *20*, 116–117.

Wan, J., Oliver, V.F., Wang, G., Zhu, H., Zack, D.J., Merbs, S.L., and Qian, J. (2015). Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. BMC Genomics *16*, 49.

Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.

Wang, L., Zhang, J., Duan, J., Gao, X., Zhu, W., Lu, X., Yang, L., Zhang, J., Li, G., Ci, W., et al. (2014). Programming and inheritance of parental DNA methylomes in mammals. Cell *157*, 979–991.

Wang, Z., Schones, D.E., and Zhao, K. (2009). Characterization of human epigenomes. Current Opinion in Genetics & Development *19*, 127.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Peng, W., Zhang, M.Q., et al. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903.

Wanrooij, S., Fusté, J.M., Farge, G., Shi, Y., Gustafsson, C.M., and Falkenberg, M. (2008). Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. Proc Natl Acad Sci U S A *105*, 11122–11127.

Waterland, R.A., and Jirtle, R.L. (2003). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. Mol. Cell. Biol. *23*, 5293–5300.

Waterland, R.A. (2014). Epigenetic mechanisms affecting regulation of energy balance: many questions, few answers. Annu. Rev. Nutr. *34*, 337–355.

Waterland, R.A., Dolinoy, D.C., Lin, J.-R., Smith, C.A., Shi, X., and Tahiliani, K.G. (2006). Maternal methyl supplements increase offspring DNA methylation atAxin fused. Genesis *44*, 401–406.

Waterland, R.A., Travisano, M., and Tahiliani, K.G. (2007). Diet-induced hypermethylation at agouti viable yellow is not inherited transgenerationally through the female. Faseb J. *21*, 3380–3385.

Wei, J., Lin, Y., Li, Y., Ying, C., Chen, J., Song, L., Zhou, Z., Lv, Z., Xia, W., Chen, X., et al. (2011). Perinatal Exposure to Bisphenol A at Reference Dose Predisposes Offspring to Metabolic Syndrome in Adult Rats on a High-Fat Diet. Endocrinology *152*, 3049–3061.

Weinhouse, C., Anderson, O.S., Bergin, I.L., Vandenbergh, D.J., Gyekis, J.P., Dingman, M.A., Yang, J., and Dolinoy, D.C. (2014). Dose-Dependent Incidence of Hepatic Tumors in Adult Mice following Perinatal Exposure to Bisphenol A. Environ Health Perspect 1–7.

Weksberg, R. (2010). Imprinted genes and human disease. Am. J. Med. Genet. *154C*, 317–320.

Whitelaw, N.C., and Whitelaw, E. (2008). Transgenerational epigenetic inheritance in health and disease. Current Opinion in Genetics & Development *18*, 273–279.

Wolff, G.L. (1971). Genetic modification of homeostatic regulation in the mouse. American Naturalist 241–252.

Wolff, G.L. (1978). Influence of maternal phenotype on metabolic differentiation of agouti locus mutants in the mouse. Genetics *88*, 529–539.

Wolff, G.L., Kodell, R.L., Moore, S.R., and Cooney, C.A. (1998). Maternal epigenetics and methyl supplements affect agouti gene expression in Avy/a mice. Faseb J. *12*, 949–957.

Wolstenholme, J.T., Edwards, M., Shetty, S.R.J., Gatewood, J.D., Taylor, J.A., Rissman, E.F., and Connelly, J.J. (2012). Gestational Exposure to Bisphenol A Produces Transgenerational Changes in Behaviors and Gene Expression. Endocrinology *153*, 3828–3838.

Wolstenholme, J.T., Goldsby, J.A., and Rissman, E.F. (2013). Transgenerational effects of prenatal bisphenol A on social recognition. Horm Behav *64*, 833–839.

Wong, R.L.Y., Wang, Q., Treviño, L.S., Bosland, M.C., Chen, J., Medvedovic, M., Prins, G.S., Kannan, K., Ho, S.-M., and Walker, C.L. (2015). Identification of secretaglobin Scgb2a1 as a target for developmental reprogramming by BPA in the rat prostate. Epigenetics *10*, 127–134.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science *329*, 444–448.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell *148*, 816–831.

Xin, F., Susiarjo, M., and Bartolomei, M.S. (2015). Multigenerational and transgenerational effects of endocrine disrupting chemicals: A role for altered epigenetic regulation? Seminars in Cell and Developmental Biology 1–10.

Xu, J.-C., Fan, J., Wang, X., Eacker, S.M., Kam, T.-I., Chen, L., Yin, X., Zhu, J., Chi, Z., Jiang, H., et al. (2016). Cultured networks of excitatory projection neurons and inhibitory interneurons for studying human cortical neurotoxicity. Science Translational Medicine *8*, 333ra48–333ra48.

Yan, H., Yuan, W., Velculescu, V.E., Vogelstein, B., and Kinzler, K.W. (2002). Allelic variation in human gene expression. Science *297*, 1143–.

Yang, A.S., Estécio, M.R.H., Doshi, K., Kondo, Y., Tajara, E.H., and Issa, J.-P.J. (2004). A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Res. *32*, e38.

Yang, X., Han, H., De Carvalho, D.D., Lay, F.D., Jones, P.A., and Liang, G. (2014). Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. Cancer Cell *26*, 577–590.

Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J.-P., Nissim-Rafinia, M., Cohen, A.-H.S., Rippe, K., Meshorer, E., et al. (2015). HP1 Is Involved in Regulating the Global Impact of DNA Methylation on Alternative Splicing. Cell Rep *10*, 1122–1134.

Zhou, R., Zhang, Z., Zhu, Y., Chen, L., and Sokabe, M. (2009). Deficits in development of synaptic plasticity in rat dorsal striatum following prenatal and neonatal exposure to low-dose bisphenol A. Neuroscience *159*, 161–171.

Zhu, J., Lee, K.P., Spencer, T.J., Biederman, J., and Bhide, P.G. (2014). Transgenerational transmission of hyperactivity in a mouse model of ADHD. J. Neurosci. *34*, 2768–2773.

Ziller, M.J., Hansen, K.D., Meissner, A., and Aryee, M.J. (2015). Coverage recommendations for

methylation analysis by whole-genome bisulfite sequencing. Nat Meth *12*, 230–2–1pfollowing232.

Zoeller, R.T., Bansal, R., and Parris, C. (2005). Bisphenol-A, an Environmental Contaminant that Acts as a Thyroid Hormone Receptor Antagonist in Vitro, Increases Serum Thyroxine, and Alters RC3/Neurogranin Expression in the Developing Rat Brain. Endocrinology *146*, 607–612.

Appendices

Appendix I: Supplemental Materials for Chapter 2

External Files

Supplemental Table S2.1 Methylation across well-characterized, imprinted (gASMs)

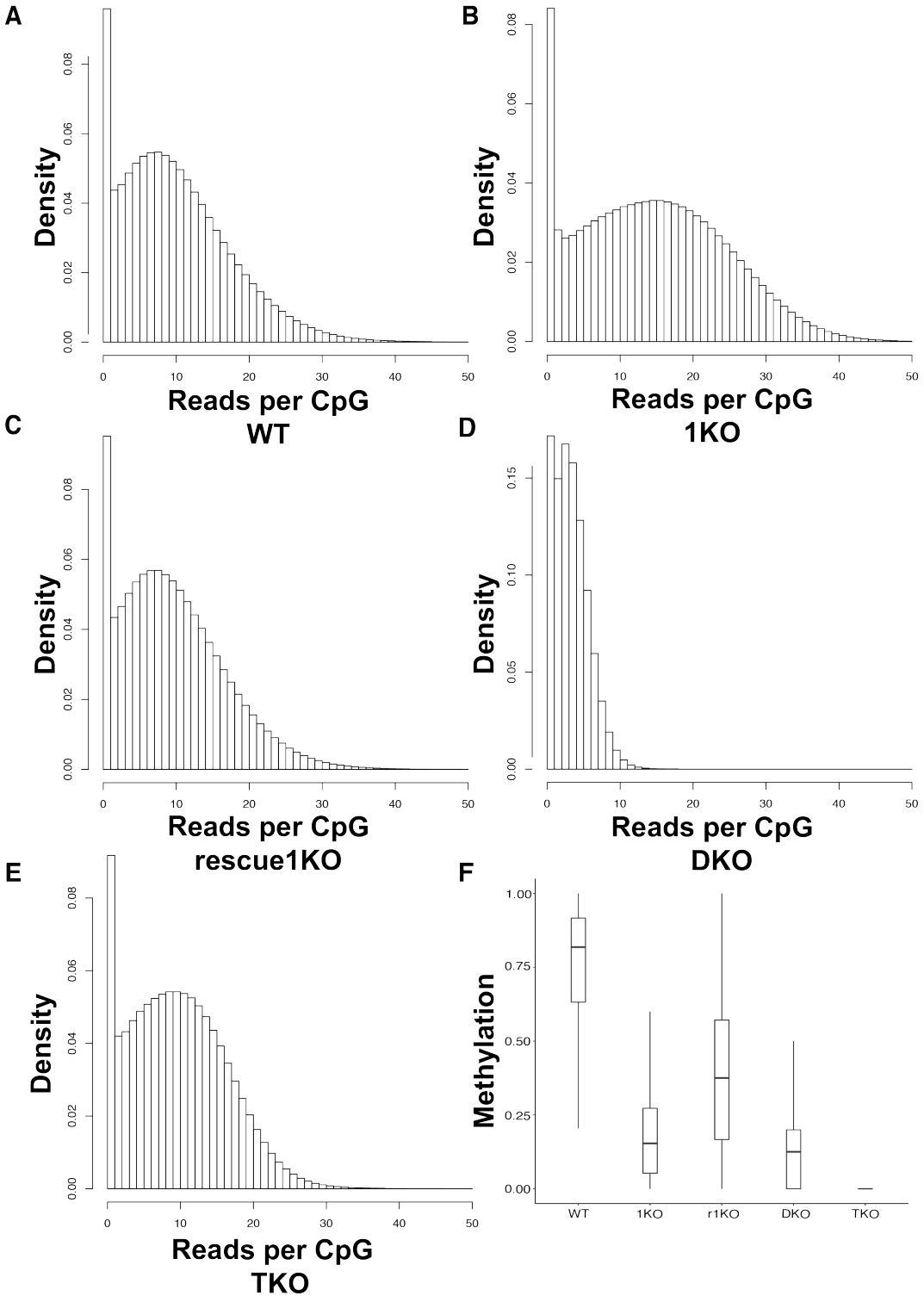
Imprinted gASM[a]	Position (mm10[b])	WT[c]	1KO[c]	DKO[c]	TKO[c]	r1KO[d]
H19 ICR	chr7:142586025-142586025	0.6718	0.0662	0.1225	0.0098	0.1389
Rasgrf1 ICR	chr9:89880290-89880290	0.6919	0.1217	0.1586	0.0010	0.2721
Gtl2/Meg3 ICR	chr12:109534883-109534883	0.4846	0.0329	0.3062	0.0039	0.0652
Gpr1	chr1:63200726-63200726	0.3686	0.0106	0.0130	0.0075	0.0223
Mcts2 (*H13*)	chr2:152687433-152687433	0.3755	0.0077	0.0029	0.0043	0.0138
Nespas-GnasXL ICR	chr2:174301063-174301063	0.4624	0.0185	0.0811	0.0066	0.0370
Gnas 1A ICR	chr2:174329137-174329137	0.1721	0.0367	0.0327	0.0051	0.0911
Peg10	chr6:4749483-4749483	0.3898	0.0088	0.0289	0.0063	0.0206
Nap1L5 (*Herc3*)	chr6:58907397-58907397	0.3486	0.0102	0.0149	0.0020	0.0174
Peg1/Mest	chr6:30739966-30739966	0.4668	0.0145	0.0041	0.0046	0.0337
Peg3 ICR	chr7:6731546-6731546	0.4374	0.0076	0.1501	0.0069	0.0082
Snrpn ICR	chr7:60006697-60006697	0.3590	0.0056	0.0302	0.0038	0.0079
Inpp5f_v2 (*Inpp5f*)	chr7:128688735-128688735	0.4479	0.0165	0.1271	0.0007	0.0511
Kcnq1ot1 ICR	chr7:143296905-143296905	0.3931	0.0052	0.0042	0.0030	0.0051
Plagl1	chr10:13092325-13092325	0.5321	0.0347	0.0952	0.0086	0.0575
Grb10 ICR	chr11:12027097-12027097	0.3187	0.0098	0.0047	0.0050	0.0286
Zrsr1 (*Commd1/Murr1*)	chr11:22974212-22974212	0.4402	0.0065	0.0060	0.0046	0.0092
Peg13/*Trappc9*	chr15:72810912-72810912	0.5334	0.0096	0.1620	0.0023	0.0153
Slc38a4	chr15:97055299-97055299	0.0112	0.0294	0.0242	0.0133	0.0216
Igf2r ICR	chr17:12742950-12742950	0.4472	0.0154	0.0386	0.0057	0.0259
Impact	chr18:12975001-12975001	0.4133	0.0075	0.0029	0.0043	0.0022

[a] Imprinted germline allele-specific methylation (gASM; also referred to as imprinted germline DMRs) in common from two sources (Arnaud, 2010 PMID: 20501788; MacDonald and Mann, 2013 PMID: 23893518)

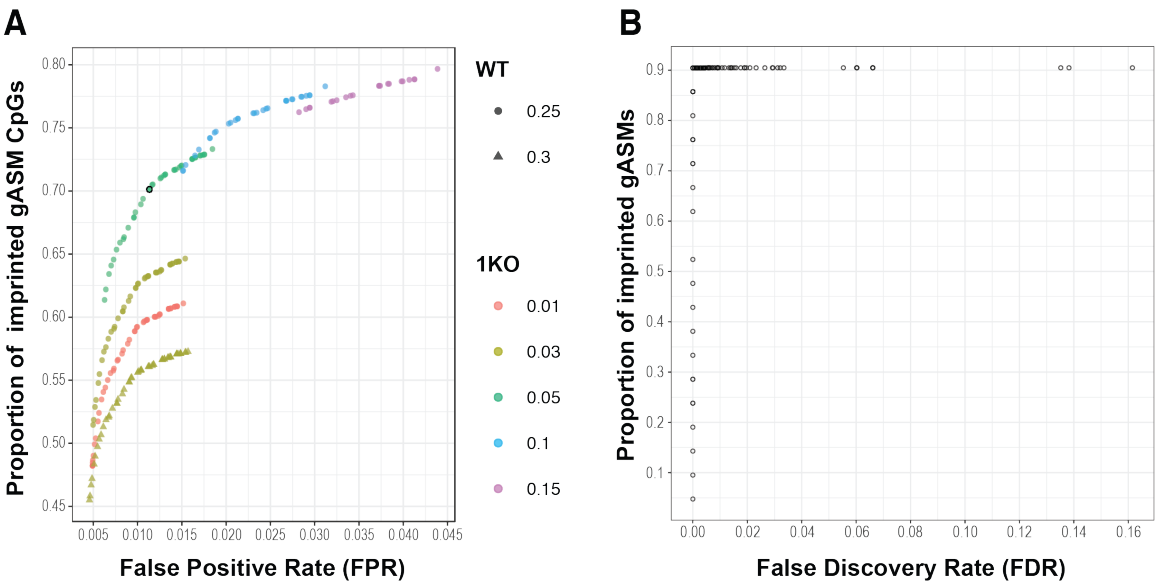
[b] Positions from mm9 for gASMs (MacDonald and Mann, 2013 PMID: 23893518) converted to UCSC mm10 with LiftOver from UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgLiftOver; Kent, 2002 PMID: 12045153)

[c] Methylation at gASMs for wild type (WT, J1) and DNMT-deficient ESCs including the loss of maintenance DNMT1 (*Dnmt1*$^{-/-}$, 1KO), of *de novo* DNMT3a/3b (*Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$, DKO), and of all three (*Dnmt1*$^{-/-}$/*Dnmt3a*$^{-/-}$/*Dnmt3b*$^{-/-}$, TKO). Average methylation of imprinted gASMs is significantly lower in 1KO (0.023), DKO (0.067), and TKO (0.005) compared to WT (0.417, $p = 1.1 \times 10^{-5}$). Compared to TKO, gASMs have significantly higher methylation in both 1KO ($p = 5.7 \times 10^{-6}$) and DKO ($p = 6.3 \times 10^{-4}$). Difference in methylation of gASMs between 1KO and DKO is not significant ($p = 0.085$). P-values determined by pairwise Wilcoxon rank-sum tests (paired) and adjusted for multiple comparisons by Bonferroni method.

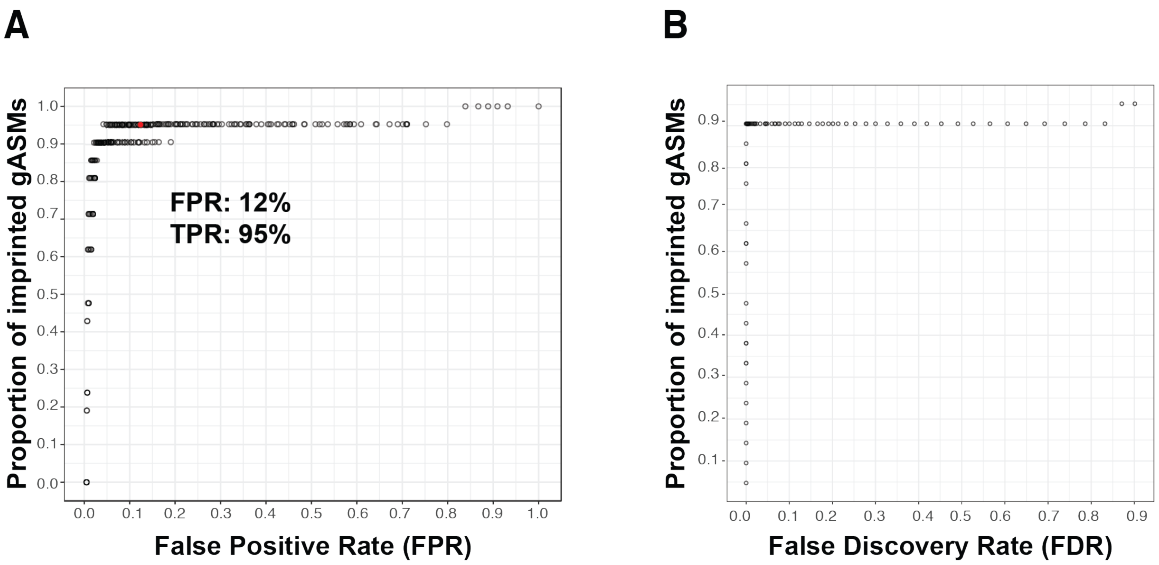
[d] Average methylation at gASMs for DNMT1-rescued 1KO cells (*Dnmt1*$^{-/-}$ + *Dnmt1* cDNA, r1KO) is 0.04



Supplemental Figure S2.1 Coverage and methylation level for WT, 1KO, r1KO, DKO, and TKO embryonic stem cells (ESCs). (A–E) Coverage per CpG site in WT **(A)**, 1KO **(B)**, r1KO **(C)**, DKO **(D)**, and TKO **(E)** ESCs. **(F)** Distribution of CpG methylation for CpG sites with minimum coverage of five in ESCs. Global average methylation for all covered CpG sites in WT, 1KO, r1KO, DKO, and TKO was 0.727, 0.176, 0.369, 0.157, 0.006, respectively.

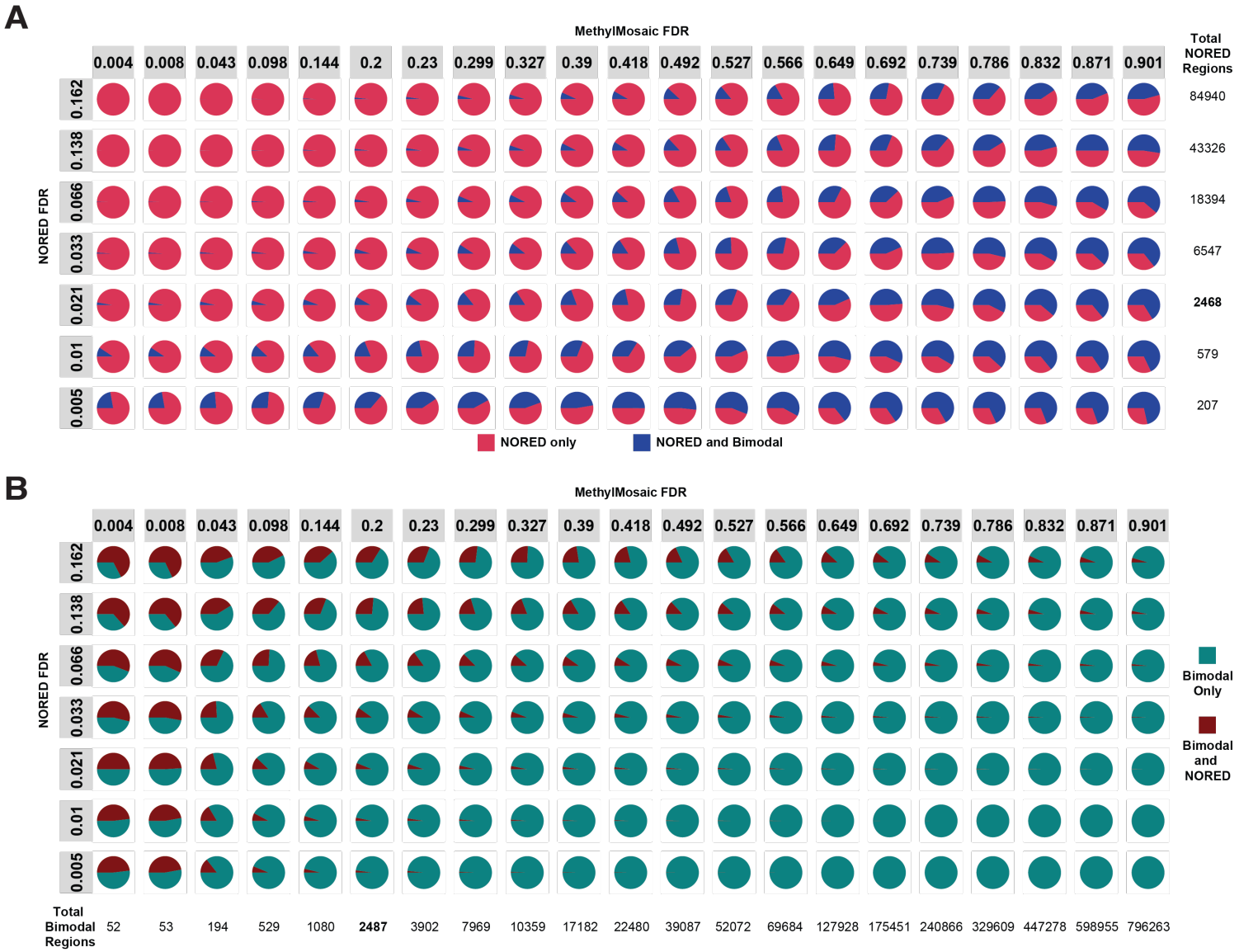


Supplemental Figure S2.2 NORED method evaluation. (A) ROC-like curve for individual CpG sites that meet three NORED criteria at various cutoffs for WT, 1KO, and r1KO. Cutoff for WT parameter is a circle for 0.25 and a triangle for 0.30. Cutoff for 1KO parameter is indicated by color. Each circle/triangle represents a tested r1KO value for combinations of WT and 1KO. For r1KO values are between WT and 1KO cutoffs at increments of 0.005. Circle with black outline indicates values for three parameters to select individual CpG sites to combine into regions (r1KO = 0.125). **(B)** Proportion of gASMs identified compared to region-level FDR

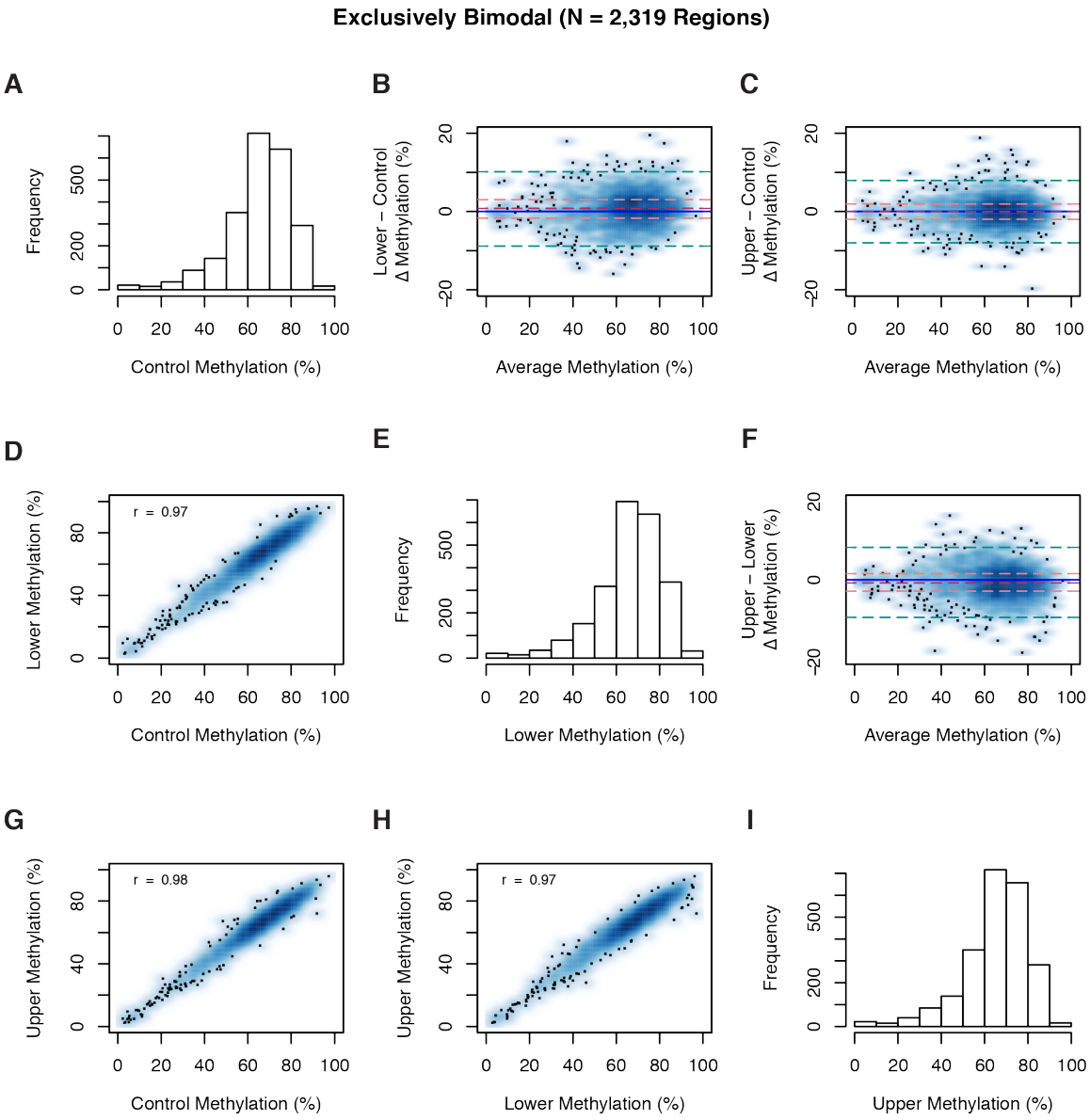


Supplemental Figure S2.3 MethylMosaic method evaluation. (A) ROC-like curve for CpG-centered windows at various cutoffs for hypermethylation and hypomethylation indices (see Methods) Red point indicates bounds for indices (0.2–0.75) to select CpG-centered windows to combine into regions. **(B)** Proportion of gASMs identified at region-level FDR.

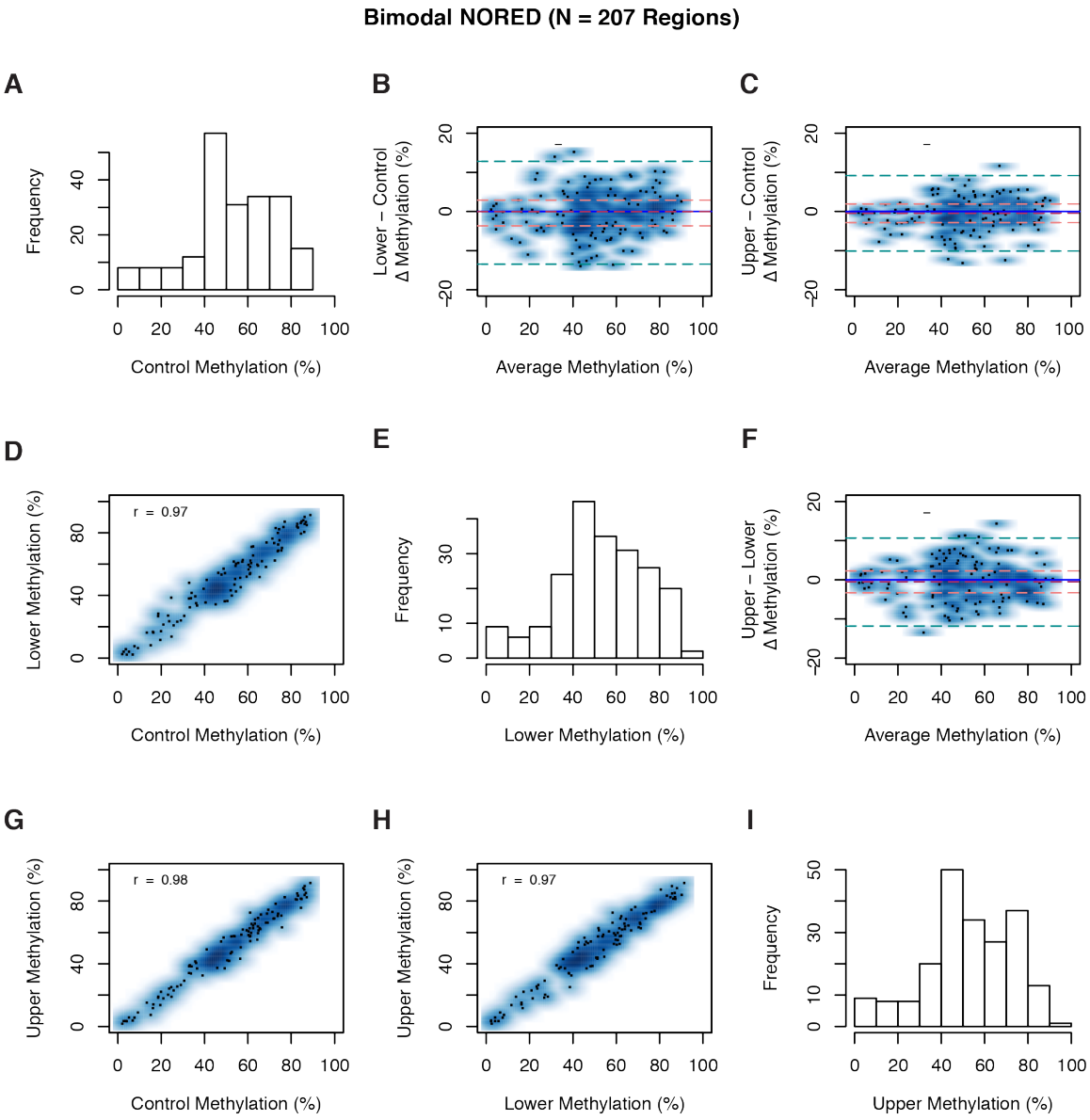
Supplemental Figure S2.4 Overlap of NORED and MethylMosaic demonstrates that two methods are independent, but complementary. (A) Proportion of NORED regions that are also bimodal in MethylMosaic at various FDRs for each method. Total number of NORED regions at each FDR cutoff is listed to the right. **(B)** Proportion of MethylMosaic regions that are also NORED at various FDRs for each method. Total number of MethylMosaic regions at each FDR cutoff is listed below.


A
MethylMosaic FDR
0.004
0.008
0.043
0.098
0.144
0.2
0.23
0.299
0.327
0.39
0.418
0.492
0.527
0.566
0.649
0.692
0.739
0.786
0.832
0.871
0.901
Total NORED Regions
NORED FDR
0.162
0.138
0.066
0.033
0.021
0.01
0.005
84940
43326
18394
6547
2468
579
207
NORED only
NORED and Bimodal
B
MethylMosaic FDR
Bimodal Only
Bimodal and NORED
Total Bimodal Regions
52
53
194
529
1080
2487
3902
7969
10359
17182
22480
39087
52072
69684
127928
175451
240866
329609
447278
598955
796263

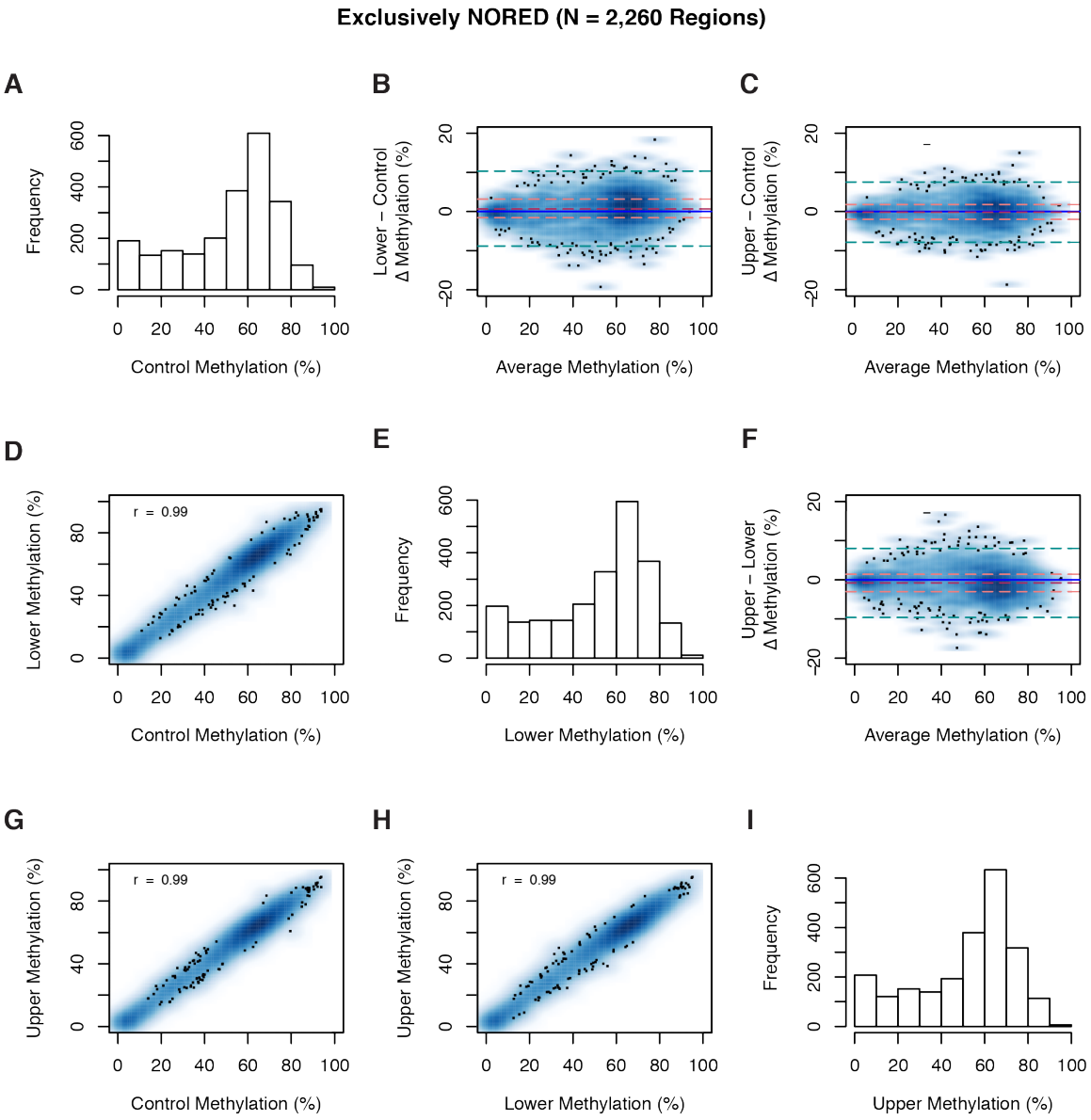
Appendix II: Supplemental Materials for Chapter 3



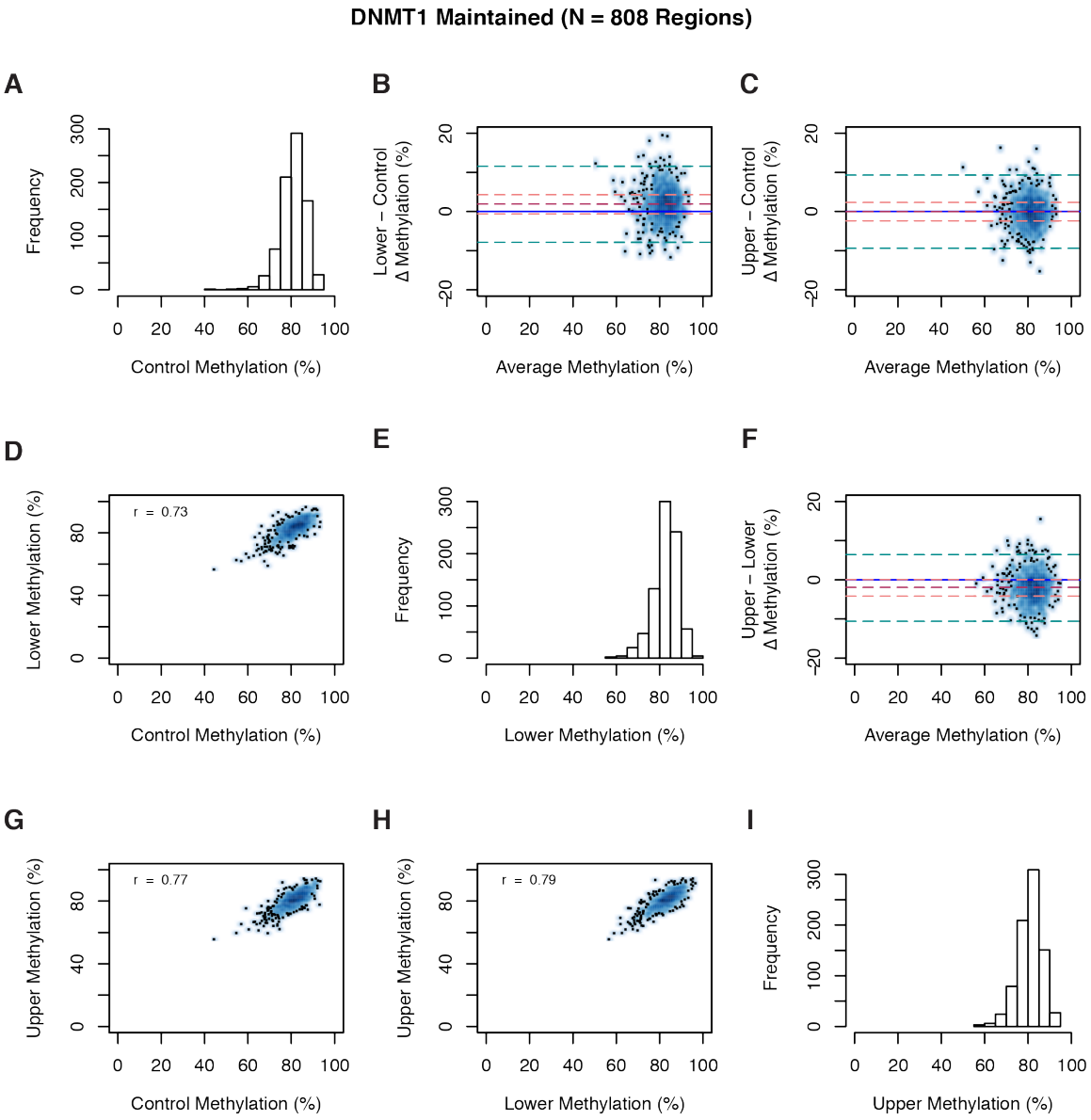
Supplemental Figure S3.1Characterization of region-level methylation within bimodal regions in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for bimodal regions from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at bimodal regions. Average methylation (x-axis) indicates the average methylation at each region for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 0.74%, -0.024%, and -0.78% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the regions. IQR (orange dashed line) was 4.47%, 3.96%, and 4.48% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at bimodal regions for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



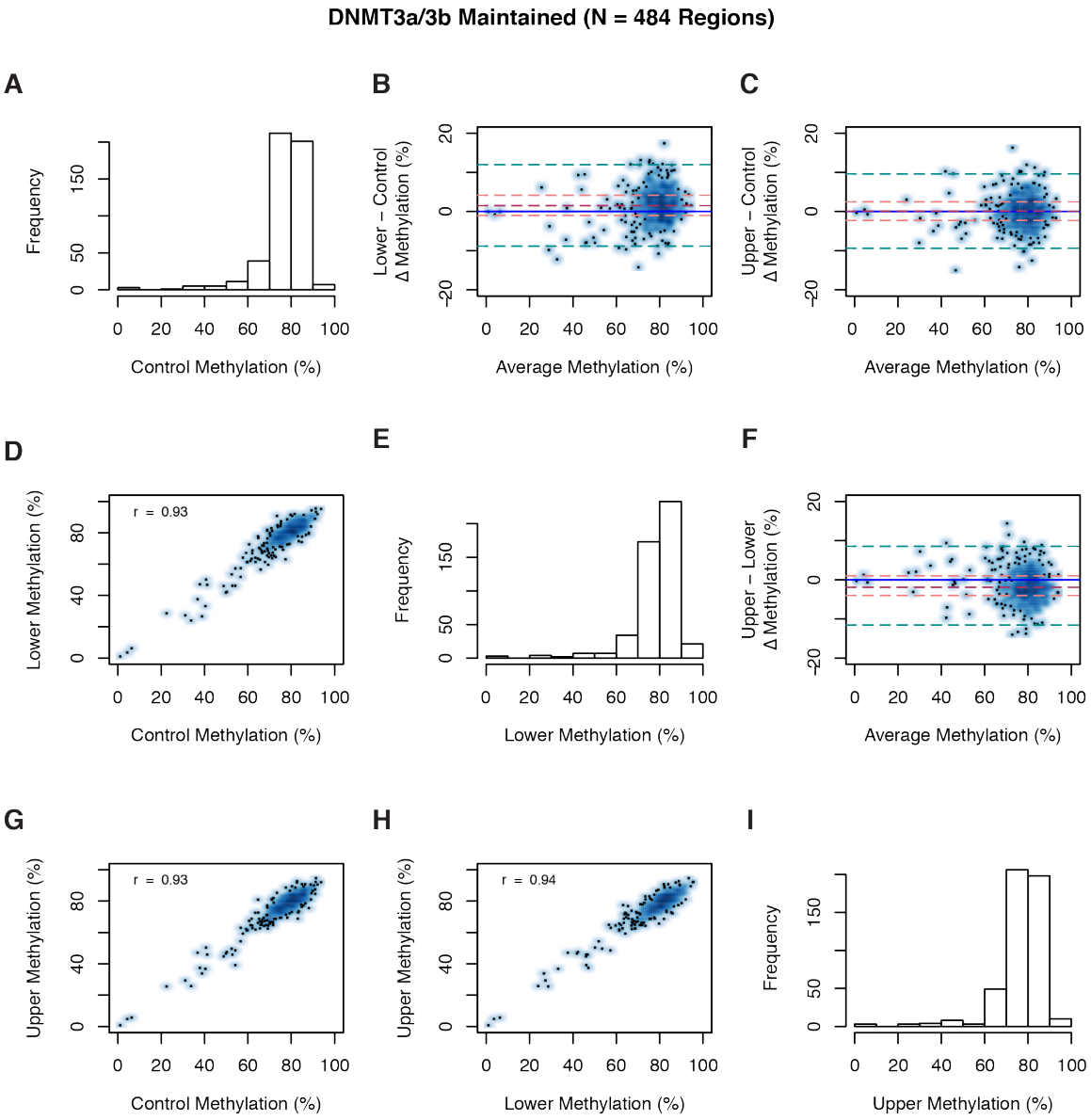
Supplemental Figure S3.2 Characterization of region-level methylation within bimodal NORED regions in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for bimodal NORED regions from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at bimodal NORED regions. Average methylation (x-axis) indicates the average methylation at each region for two dose groups within the comparison. Median (maroon dashed line) methylation difference was -0.088%, -0.44%, and -0.49% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the regions. IQR (orange dashed line) was 6.57%, 4.82%, and 5.64% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at bimodal NORED regions for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



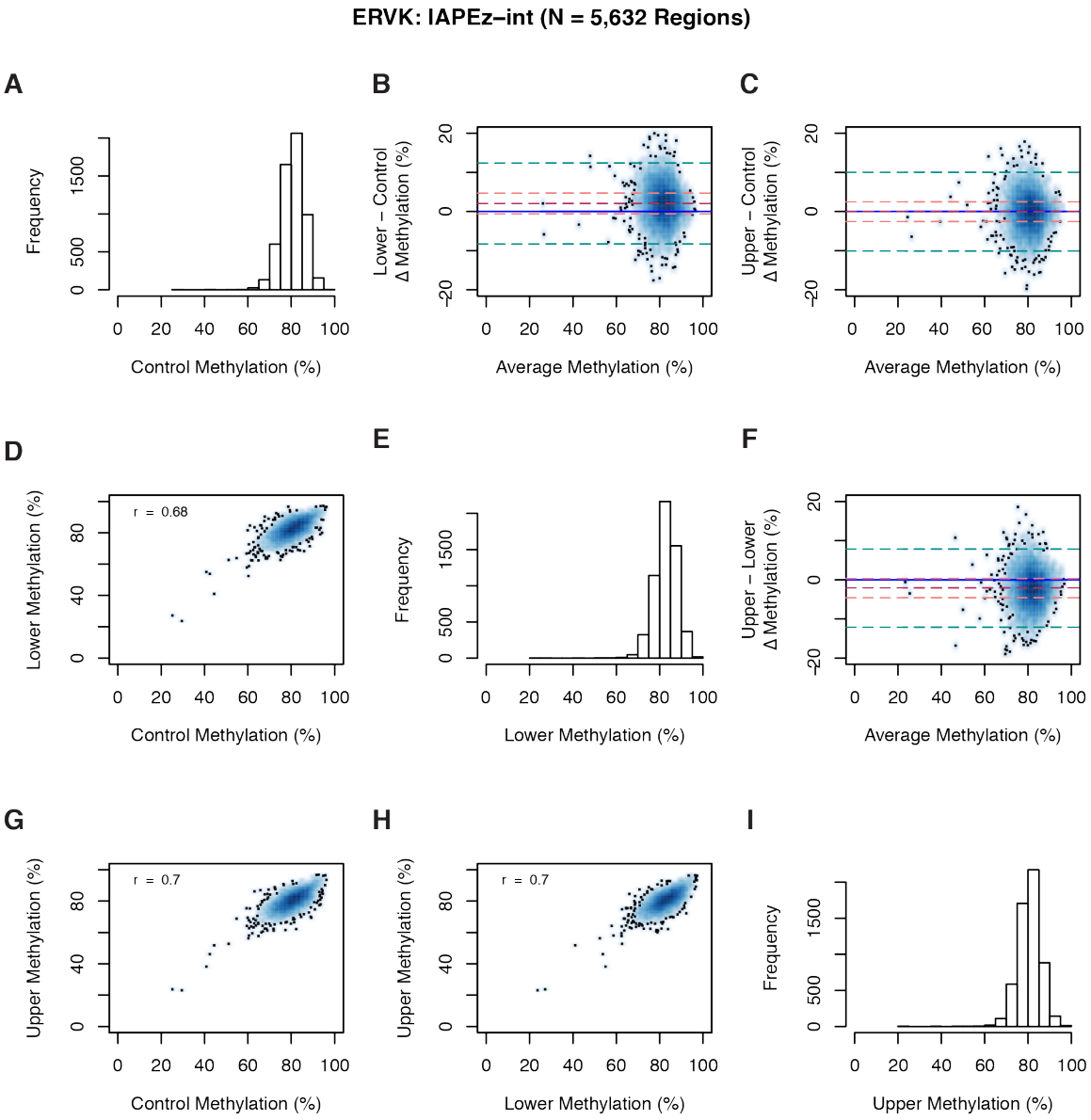
Supplemental Figure S3.3 Characterization of region-level methylation within NORED regions in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for NORED regions from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at NORED regions. Average methylation (x-axis) indicates the average methylation at each region for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 0.62%, -0.23%, and -0.77% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the regions. IQR (orange dashed line) was 4.79%, 3.83%, and 4.42% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at NORED regions for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



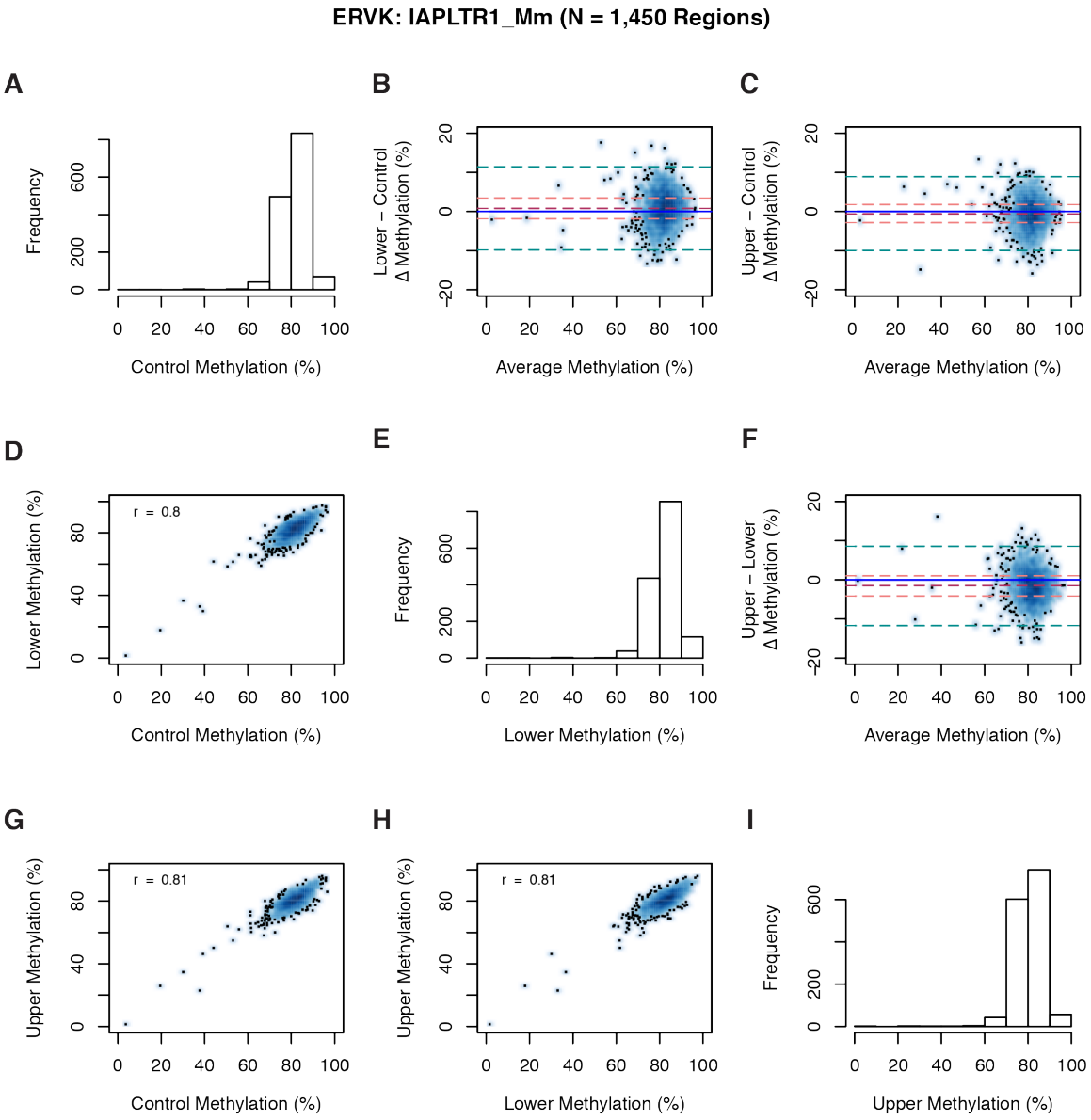
Supplemental Figure S3.4 Characterization of region-level methylation within DNMT1-maintained regions in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for DNMT1-maintained regions from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at DNMT1-maintained regions. Average methylation (x-axis) indicates the average methylation at each region for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 1.9%, -0.076%, and -2.0% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the regions. IQR (orange dashed line) was 4.85%, 4.68%, and 4.26% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at DNMT1-maintained regions for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



Supplemental Figure S3.5 Characterization of region-level methylation within DNMT3a/3b-maintained regions in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for DNMT3a/3b-maintained regions from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at DNMT3a/3b-maintained regions. Average methylation (x-axis) indicates the average methylation at each region for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 1.6%, -0.16%, and -1.9% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the regions. IQR (orange dashed line) was 5.20%, 4.76%, and 5.02% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at DNMT3a/3b-maintained regions for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



Supplemental Figure S3.6 Characterization of region-level methylation within IAPEz retrotransposons in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for IAPEz elements from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at IAPEz elements. Average methylation (x-axis) indicates the average methylation at each element for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 2.1%, 2.7 x 10^{-4}%, and -2.1% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the elements. IQR (orange dashed line) was 5.20%, 5.05%, and 4.97% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at IAPEz elements for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**.



Supplemental Figure S3.7 Characterization of region-level methylation within IAPLTR1 Mm retrotransposons in BPA-exposed and control fetal livers. (A, E, I) Region-level methylation for IAPLTR1 Mm elements from three control **(A)**, four lower **(E)**, and four upper **(I)** dose replicates. **(B, C, F)** Bland-Altman plots showing distribution of region-level methylation differences (Δ) between lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)** dose groups at IAPLTR1 Mm elements. Average methylation (x-axis) indicates the average methylation at each element for two dose groups within the comparison. Median (maroon dashed line) methylation difference was 0.85%, -0.63%, and -1.5% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Interquartile range (IQR) represents the middle 50 percent of the data as it is calculated as the difference between 75th percentile and the 25th percentile change in methylation for the elements. IQR (orange dashed line) was 5.31%, 4.68%, and 5.08% for lower and control **(B)**, upper and control **(C)**, and upper and lower **(F)**, respectively. Cyan dashed line is at 1.5 x IQR. **(D, G, H)** Correlation (Pearson's r) for region-level methylation at IAPLTR1 Mm elements for three comparisons: lower vs. control **(D),** upper vs. control **(G)**, and upper vs. lower **(H)**

Appendix III: Supplemental Materials for Chapter 4

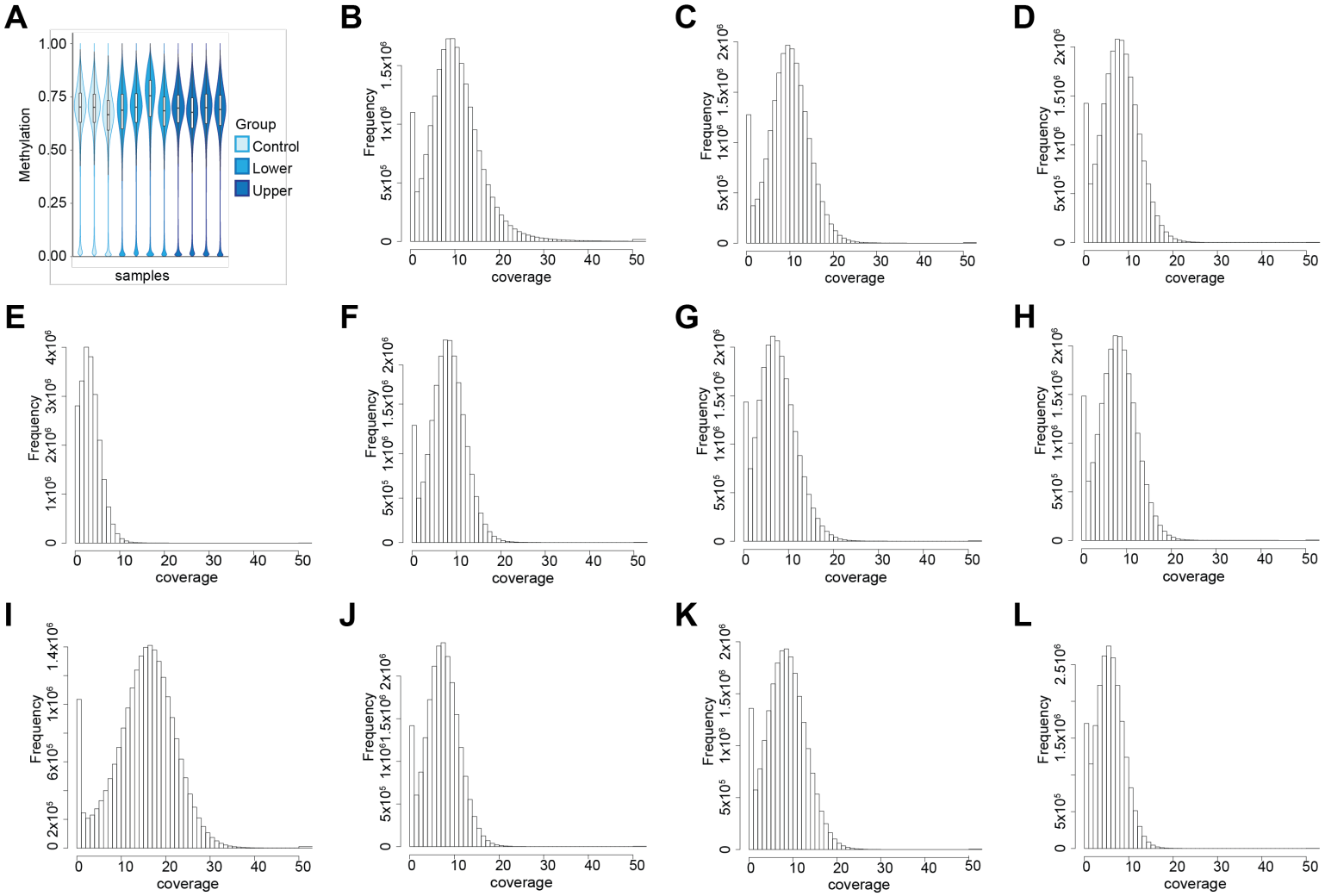
External Files

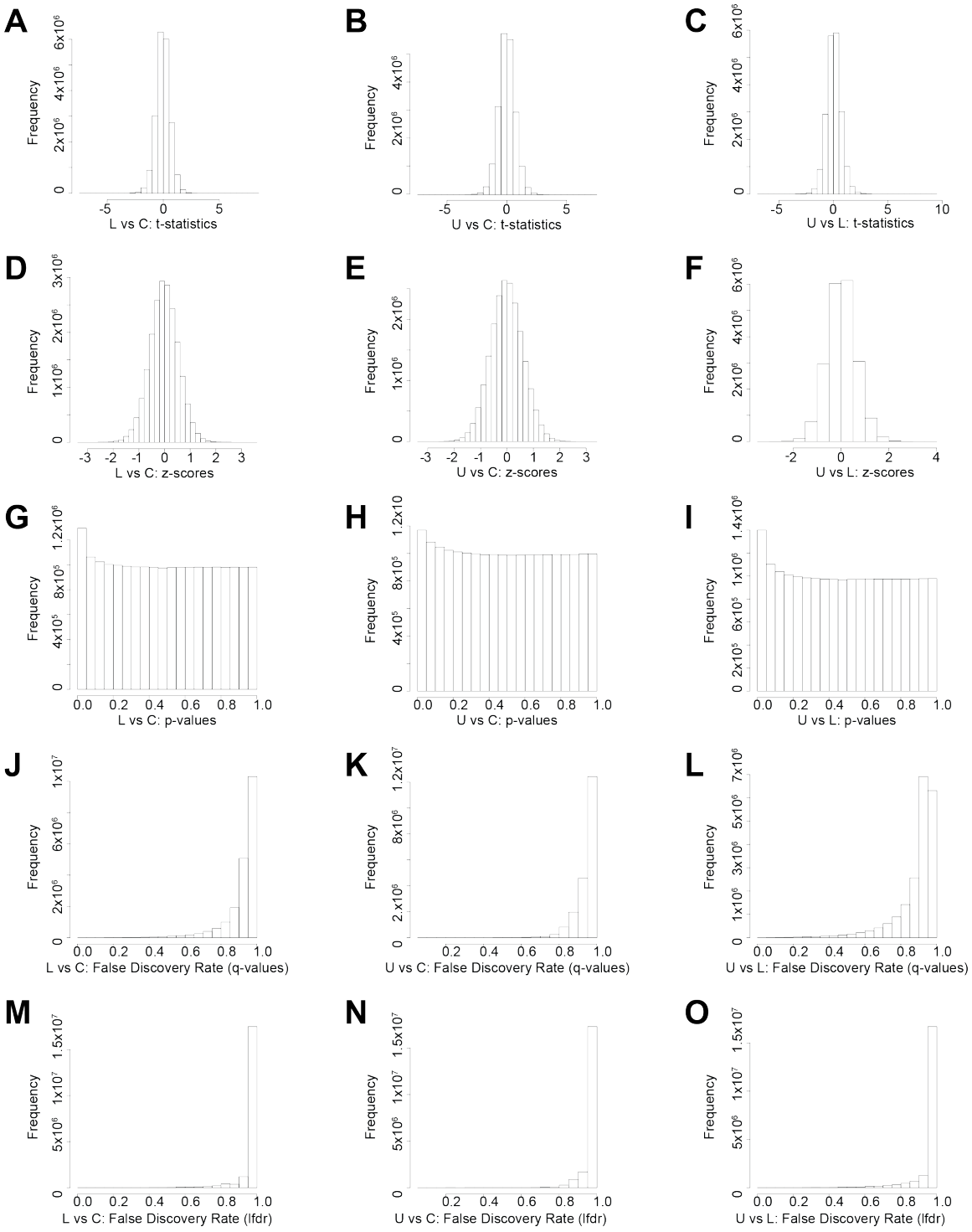
Supplemental Table S4.1 Alignment metrics for RNA-seq data.

Sample	Reads	# of reads	%Total aligned	% Abundant	% Unaligned	Median CV Coverage Uniformity
Control1	75/75	50236,339	90.76%	12.46%	9.24%	0.99
Control2	75/75	48,427,199	90.07%	14.31%	9.93%	0.99
Control3	75/75	38,842,843	90.84%	17.03%	9.16%	0.99
Lower1	75/75	50,321,213	89.35%	8.74%	10.65%	0.94
Lower2	75/75	46,876,155	91.69%	13.68%	8.31%	1.02
Lower3	75/75	42,534,926	90.18%	8.95%	9.82%	1.04
Upper1	75/75	46,926,847	91.34%	15.17%	8.66%	1.14
Upper2	75/75	35,930,488	94.01%	17.74%	5.99%	1.16
Upper3	75/75	72,909,842	93.96%	17.74%	6.04%	1.1

Reads = number and length of reads (bp)
# of reads = total number of reads passing filter
% total aligned = percentage of reads passing filter that aligned to the reference including abundant reads
% abundant = percentage of reads that align to abundant transcripts such as mitochondrial and ribosomal sequences
% unaligned = percentage of reads that do not align with reference sequences
median CV coverage uniformity = the median coefficient of variation of coverage of the 1000 most highly expressed transcripts, as reported by the CollectRNASeqMetrics utility from Picards tool

Supplemental Figure S4.1 Genome-wide coverage and methylation for WGBS samples. (A) Violin and box plots of CpG methylation for control (Ca, Cb, and Cc), lower (La, Lb, Lc, and Ld), and upper dose (Ua, Ub, Uc, Ud) male fetal livers. **(B–L)** Sequencing reads coverage of autosomal and sex chromosome CpG sites for WGBS samples. Control dose replicates: Ca, Cb, and Cc **(B–D**). Lower BPA dose replicates La, Lb, Lc, and Ld **(E–H)**. Upper BPA dose replicates: Ua, Ub, Uc, Ud **(I–L)**.


A
Methylation
1.00
0.75
0.50
0.25
0.00
samples
Group
Control
Lower
Upper
B
C
D
E
F
G
H
I
J
K
L
Frequency
coverage
0
10
20
30
40
50



Supplemental Figure S4.2 Distribution of statistics for single-CpG methylation differences in BPA-exposed male fetal livers. (A–O) Lower vs. control (L vs C), upper vs. control (U vs C), and upper vs lower (U vs L) refer to pairwise comparisons between control (n = 3), lower (n = 4), and upper (n = 4) dose groups. **(A–C)** Bsmooth bsseq (R package) t-statistics. **(D–F)** T-statistics transformed to z-scores. **(G–I)** P-values calculated from z-scores. **(J–L)** Tail-area-based false discovery rate (q-value). **(M–O)** Empirical null-based false discovery rate (lfdr).

Curriculum Vitae

SUZANNE N. MARTOS

smartos1@jhu.edu

SUMMARY OF QUALIFICATIONS

Epigenetics Ph.D. candidate with laboratory and computational experience in

- Targeted and genome-wide approaches for characterizing epigenetic marks.
- Cell culture of primary stem cells, embryonic stem cells, and differentiation of neural progenitor cells to neurons.
- Identification of epigenetic changes from genetic (e.g., gene knockout) and chemical (e.g., toxicant treatment) perturbations.
- Development of novel bioinformatics approaches to address complex epigenetic questions.

EDUCATION

Johns Hopkins Bloomberg School of Public Health Baltimore, MD

Ph.D. in Environmental Health and Engineering: Toxicology, Physiology & Molecular Mechanisms Expected August 2017

- GPA: 4.0
- Additional Coursework: Bioinformatics, Statistical Computing, Statistics for Genomics, Molecular Biology and Genomics, Pathways and Regulation, Cell Structure and Dynamics, Macromolecular Structure Analysis

M.H.S in Environmental Health: Toxicology and Pathophysiology May 2010

- GPA: 4.0
- Certificate: Risk Sciences and Public Policy

Cornell University Ithaca, NY

B.S. in Biological Sciences; B.S. in Science of Earth Systems: Biogeochemistry May 2007

- Meinig Family Cornell National Scholars: selected based on leadership, academic excellence and service to community (< 2% of undergraduate population)
- Earth and Environmental Systems Field Program in Hawaii
- Founding member of Alpha Beta Chapter of Alpha Xi Delta
- Cornell Women's Rugby Football Club
- Additional Coursework: Introductory Computer Programming for Biological and Environmental Engineers, Honors Calculus II, Linear Algebra and Differential Equations

RESEARCH EXPERIENCE

Johns Hopkins University Baltimore, MD

Graduate Researcher with Dr. Zhibin Wang April 2013–present

Conducted thesis research in the Wang Laboratory of Human Environmental Epigenomes to determine how the plasticity and heritability of epigenetic marks, such as DNA methylation, contribute to mechanisms of toxicity from prenatal exposures.

- Collaborated with Dr. Marisa Bartolomei (University of Pennsylvania) and Dr. Martha Susiarjo (University of Rochester) to characterize DNA methylation and gene expression changes in fetal livers of male mice exposed in utero to Bisphenol A (BPA).
- Developed a permutation-based approach to identify non-restorable DNA methylation loss (Non-Rescued Differentially Methylated Region, NORED) genome-wide in mouse embryonic stems cells.
- Developed algorithms to quantify read-level DNA methylation and implement a permutation-based approach to identify bimodal methylation patterns throughout the mouse genome (MethylMosaic).
- Developed data analysis pipelines combining publicly available software (Bowtie, Bismark, MACS and R packages) and custom Bash and R scripts for use on a local Linux server or SLURM-managed high-performance computing cluster to process and analyze data from whole-genome bisulfite sequencing (DNA methylation), chromatin immunoprecipitation (ChIP) sequencing (histone modifications), nucleosome positioning sequencing and mRNA sequencing experiments for mouse, cow, and/or human genomes.

- Constructed libraries for whole-genome bisulfite sequencing, nucleosome sequencing and RNA sequencing.
- Prepared bisulfite PCR products and analyzed data for targeted bisulfite amplicon and Sanger sequencing to characterize DNA methylation.
- Cultured mouse embryonic stem cells and treated with environmental toxicants.
- Extracted DNA and RNA from mouse embryonic stem cells and tissues.

Johns Hopkins University Baltimore, MD
Graduate Rotation with Dr. Kathleen Gabrielson December 2012–March 2013

Conducted rotation project in the Department of Molecular and Comparative Pathobiology in (School of Medicine) to investigate effects of neuregulin-1β on ErbB2 levels and combined toxicity of Doxorubicin and Herceptin.

- Cultured human mesenchymal stem cells (primary cells), pretreated with neuregulin-1β, and treated with Doxorubicin and Herceptin alone or in combination.
- Evaluated cell viability using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) and lactate dehydrogenase (LDH) assays.
- Extracted protein and performed western blots.

Johns Hopkins University Baltimore, MD
Graduate Rotation with Dr. Thomas Hartung and Dr. Lena Smirnova September 2012–November 2012

Conducted rotation project in the Center for Alternatives to Animal Testing (CAAT) to analyze miRNA expression in neuronal precursor cells exposed to developmental neurotoxicants during differentiation to neurons.

- Cultured Lund Human Mesencephalic (LUHMES) cells, differentiated to neurons and treated with lead chloride or valproic acid.
- Extracted RNA and performed quantitative RT-PCR for mRNA and miRNA expression.

Columbia University Palisades, NY
Undergraduate Researcher with Dr. Dallas Abbott June 2006–August 2006

Conducted research at the Lamont-Doherty Earth Observatory of Columbia University through the National Science Foundation Research Experience for Undergraduates program to determine that the composition and distribution of impact ejecta recovered from deep-sea cores surrounding two crater candidates identified by bathymetry was consistent with an extraterrestrial impact using microscopy and X-ray fluorescence.

The Nature Conservancy Kamuela, HI
Internship Experience in Earth and Atmospheric Sciences with Dr. Jan Ostman-Lind May 2006

Collected field samples to study the infection and emergence of the Erythrina gall wasp (*Quadrastichus erythrinae*) from the Wiliwili tree (*Erythrina sandwicensis*) in dryland forests of Hawaii.

Cornell University Ithaca, NY
Undergraduate Research Assistant with Dr. Kevin Nixon August 2004–December 2005

Identified Cretaceous angiosperm flower fossils to improve the study of plant evolution and biodiversity.

TEACHING & MENTORING

Society of Toxicology Annual Meeting Baltimore, MD
Peer Mentor March 2017

Mentored undergraduate students for the Sunday Undergraduate Program

Society of Toxicology Annual Meeting New Orleans, LA
Peer Mentor March 2016

Mentored five undergraduate students for the Committee on Diversity Initiatives Undergraduate Diversity Program.

Johns Hopkins University Baltimore, MD
Teaching Assistant
Toxicological Pathology: graduate course (187.620.01) Spring 2013

Public Health Toxicology: graduate course (187.610.01) August–October 2014
Public Health Toxicology: online graduate course (187.610.81) October–December 2014

Kaplan Ithaca, NY; Vienna, VA; Arlington, Virginia
Test Prep Instructor February 2006–June 2008

Taught classes of 5 to 20 students and tutored individual students to prepare them for the Medical College Admission Test (MCAT), the Dental Admission Test (DAT), the Optometry Admission Test (OAT) and the Pharmacy College Admission Test (PCAT).

Cornell University Ithaca, NY
Teaching Assistant August 2006–May 2007

Taught labs and sections for Introductory Biology (undergraduate course: BIOG 1105 & 1106). Tutored students, graded written tests and wrote and administered oral exams to assess proficiency in material.

GOVERNMENT & CONSULTING

U.S. Environmental Protection Agency Arlington, VA
Association of Schools of Public Health Fellow September 2010–August 2012

Characterized health hazards of emerging contaminants for the Integrated Risk Information System (IRIS).

U.S. Food and Drug Administration Arlington, VA
Oak Ridge Institute for Science and Education (ORISE) Intern June 2010–August 2010

Developed recommendations for risk-based models for scenarios requiring consistent, transparent, and evidence-based decisions. Reviewed literature on multi-criteria decision analysis models focused on model development for use in risk assessment and risk management. Evaluated agency decision-making algorithms used to prioritize food safety inspections. Assessed strengths and limitations of each model and provided recommendations to expand current models by incorporating additional data and risk information.

ICF International Fairfax, VA
Research Assistant May 2008–August 2009

Collected and analyzed data, reviewed literature, solved quantitative and qualitative problems, drafted reports and presentations, developed recommendations and corresponded with clients and subcontractors. Reviewed laboratory reports and drafted Data Evaluation Record reports to determine the adequacy of toxicity studies conducted for pesticides and other chemical products for the U.S. Environmental Protection Agency (EPA). Edited content of three Environmental Impact Statements (EISs). Reviewed climate change literature for the *Corporate Average Fuel Economy (CAFE) Standards for Model Years 2011-2015* EIS. Critically evaluated EIS documents for inconsistencies, procedural errors and scientific inaccuracies to provide litigation support for transportation projects. Facilitated meetings of the National Environmental Justice Advisory Council for the U.S. EPA Office of Environmental Justice.

LEADERSHIP & SERVICE

Gordon Research Seminar
Discussion Leader, Cellular & Molecular Mechanisms of Toxicity August 2017

Peer Review
Reviewer, Environmental Research (Elsevier Publishing Company) August 2016–present
Reviewer, Book Chapter in Epigenetics in Human Disease July 2017

Society of Toxicology
Chair, Graduate Student Leadership Committee May 2016–April 2017
Programming Subcommittee Chair, Graduate Student Leadership Committee May 2015–April 2016
Member, Graduate Student Leadership Committee May 2014–April 2015
Graduate Student Representative, Committee on Diversity Initiatives May 2014–April 2015
Graduate Student Representative, National Capital Area Chapter May 2014–April 2015
Graduate Student Vice-Representative, National Capital Area Chapter May 2013–April 2014

Johns Hopkins University
Student Representative, Environmental Health Sciences Research Committee September 2014–June 2016
Board Member, Environmental Health Sciences Student Organization September 2015–June 2016

PROFESSIONAL AFFILIATIONS

Society of Toxicology
Specialty Sections: Mechanisms; Molecular and Systems Biology; In Vitro and Alternative Methods
Special Interest Groups: Women in Toxicology; Hispanic Organization of Toxicologists
Regional Chapter: National Capital Area Chapter

WORKSHOPS

Van Andel Research Institute Grand Rapids, MI
Epigenomics at VARI Workshop Attendee July 2016

Selected as one of fifteen graduate students to attend one-week intensive workshop focused on applying epigenomic methods in biomedical research to understand disease, develop therapies, and improve human health.

MANUSCRIPTS

Martos SN, Susiarjo M, Stefaniak M, Lan Y, Xin F, Wang Z., Bartolomei MS. Whole-genome bisulfite sequencing reveals epigenomic insights for Bisphenol A-induced metabolic abnormalities in mice. (*In preparation*)

Martos, SN, Li T, Ramos RB, Lou D, Dai H, Xu J, Gao G, Wang Q, An C, Zhang X, Dawson VL, Dawson TM, Ji H, Wang Z. Two approaches reveal a paradigm of 'switchable or genetics-influenced allele-specific DNA methylation (ASM)' with potential in human disease. (*In preparation*)

Martos SN, Tang W-Y, Wang Z. Elusive inheritance: Transgenerational effects and epigenetic inheritance in human environmental disease. Prog Biophys Mol Biol. Elsevier Ltd; 2015 Jul;118(1-2):44–54. PMID: 25792089

Li Z, Dai H, **Martos SN**, Xu B, Gao Y, Li T, Zhu G, Schones DE, Wang Z. Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. Genome Biol. 2015;16(1):115–30. PMID: 26032981

Willett C, Caverly Rae J, Goyak KO, Minsavage G, Westmoreland C, Andersen M, Avigan M, Duche D, Harris G, Hartung T, Jaeschke H, Kleensang A, Landesmann B, **Martos SN**, Matevia M, Toole C, Rowan A, Schultz T, Seed J, Senior J, Shah I, Subramanian K, Vinken M, Watkins P. Building shared experience to advance practical application of pathway-based toxicology: liver toxicity mode-of-action. ALTEX. 2014;31(4):500–519. PMID: 24535319

PRESENTATIONS

Martos SN, Wang Z. Identification of DNMT1-dependent regions in mouse and human embryonic stem cells provides insight into developmental epigenotoxicity. Gordon Research Conference on Cellular & Molecular Mechanisms of Toxicity, August 2017, Andover, NH (Poster)

Martos SN, Susiarjo M, Bartolomei MS, Wang Z. Whole-genome Bisulfite Sequencing Reveals Epigenetic Characteristics of Germline Imprinted Regions as Susceptible to Perturbation from Prenatal Exposure Bisphenol A. National Capital Area Chapter (NCAC) Spring Symposium, April 2017, Bethesda, MD (Poster); ***1st Place*** *Graduate Student Poster Competition*

Martos SN, Susiarjo M, Lan Y, Bartolomei MS, Wang Z. Whole-genome Bisulfite Sequencing Provides Insights Into Metabolic Dysfunction In Male Mice Exposed *In Utero* To Bisphenol A. Society of Toxicology Annual Meeting, March 2017, Baltimore, MD (**Speaker**)

Martos SN, Susiarjo M, Bartolomei MS, Wang Z. Are DNMT1-dependent Regions More Vulnerable to Epigenetic Perturbation from Bisphenol A? 25 Years of Endocrine Disruption Research: Past Lessons and Future Directions, September 2016, Bethesda, MD (Poster); ***Travel Award***

Martos SN, Susiarjo M, Bartolomei MS, Wang Z. Whole-Genome Bisulfite Sequencing Reveals Genomic Regions Vulnerable to Epigenetic Perturbation. Society of Toxicology Annual Meeting, March 2016, New Orleans, LA (Poster)

Martos SN, Susiarjo M, Bartolomei MS, Wang Z. Whole-Genome Bisulfite Sequencing Reveals Genomic Regions Vulnerable to Epigenetic Perturbation. Gordon Research Conference on Cellular & Molecular

Mechanisms of Toxicity, August 2015, Andover, NH (**Speaker** and Poster); *Carl Storm Underrepresented Minority Fellowship* ***Travel Award***

Martos SN, Sysa-Shah P, Gabrielson K. Neuregulin (NRG1) Induces a Dose-response Increase in Epidermal Growth Factor Receptor 2 (ErbB2) in hMSCs: Potential Cell-based model of ErbB2 Variability on Mechanisms of Toxicity. Society of Toxicology Annual Meeting, March 2013, San Antonio, TX (Poster); *National Capital Area Chapter of SOT Bern Schwetz Student* ***Travel Award, 1st Place***

Martos SN, Abbott D, Elkinton HD, Chivas AR, Breger D. Impact Spherules from the Craters Kanmare and Tabban in the Gulf of Carpentaria. Geological Society of America Annual Meeting, 2006, Philadelphia, PA (Poster)